Evidence-based Principles to Guide Collaborative Approaches to Evaluation: Technical Report

Research Team

Lyn M. Shulha, Queen’s University
J. Bradley Cousins University of Ottawa
Hind Al Hudib, University of Ottawa
Elizabeth Whitmore, Carleton University
Nathalie Gilbert, University of Ottawa

May, 2015

Reference as:
Acknowledgements

Correspondence: L. Shulha, Queen’s University, Faculty of Education, Duncan McArthur Hall, 511 Union Street, Kingston, ON, CANADA, K7M 5R7, lyn.shulha@queensu.ca. The authors are grateful for funding provided by the University of Ottawa. The opinions expressed in this paper do not necessarily reflect those of the sponsoring organization. The authors are indebted to Daniel Tamblyn-Watts for technical support for on-line data collection.
Abstract

This technical report introduces a set of evidence-based principles to guide evaluation practice in contexts where evaluation knowledge is collaboratively produced by evaluators and stakeholders. The data from this study evolved in four phases: two pilot phases exploring the desirability of developing a set of principles, an online questionnaire survey that drew on the expertise of practicing evaluators to identify dimensions, factors or characteristics that enhance or impede success in collaborative approaches, and finally a validation phase. The principles introduced here stem from the experiences of 320 evaluators who have engaged in collaborative approaches in a wide variety of evaluation settings and the lessons they have learned. We expect the principles to evolve over time, as evaluators learn more about collaborative approaches in context. With this in mind, we pose questions for consideration to stimulate further inquiry.

Key Words: principles, collaborative approaches, program evaluation, context, learning
Introduction and Background

The principles introduced here are an empirically derived system for thinking about collaborative approaches to evaluation. They are intended to support considerations of professional practice, both generally and in situ. In spearheading this work, we were swayed by the argument that expert behaviour is not simply a product of experience, but the disposition to use experience as a learning mechanism (Daley, 1999). The principles presented in Figure 1 are grounded in the experiences of over 300 practicing evaluators. In synthesizing what worked (or did not) during their collaborations, and in considering why these approaches worked (or did not), we began to see patterns in how evaluators’ decisions in collaborative contexts, shaped the nature of their collaborative work, and ultimately, the way they perceived success in their evaluation practices. Continuous learning about such patterns can be informative in refining what Kennedy (1983) referred to as ‘working knowledge.’

Figure 1: An Integrated Set of Principles for Use in Guiding Collaborative Approaches to Evaluation

Working knowledge is holistic and shaped by a meaningful synthesis of theory, skills, experiences, assumptions, beliefs, and one’s personal orientation to practice. A rounded working knowledge of evaluation is more likely to help practitioners function in new or unfamiliar contexts, identify complexity, and foresee the effects of potential decision options. These principles are offered as mechanisms to support the development of professional working knowledge and evaluator expertise in the use of collaborative approaches to evaluation.
Why Principles? Why Now?

To guide our work, we adopted the Oxford Dictionary (2015) definition of principles; namely, the “foundation for a system of belief or behaviour or for a chain of reasoning.” The task of generating principles that could be both explicit and well-grounded meant finding a way to reveal the nature of collaborative approaches as these exist in practice (McNiff, 2013). Rosch (1999) argued that distilling knowledge accumulated from experience is the optimal way to derive a system of principles. Patton (2015a) concurs, and suggested ways to proceed.

Principles are built from lessons that are based on evidence about how to accomplish some desired result. Qualitative inquiry is an especially productive way to generate lessons and principles precisely because purposeful sampling of information-rich cases, systematically and diligently analyzed, yield rich, contextually sensitive findings. (pp. 715-716).

In citing the work of evaluation scholars such as Cronbach and Shapiro (1978) and Stake (1983), as cited in McTaggart (1991), we are reminded that principles, particularly descriptive ones, are to be pragmatic tools that can encourage readers to extrapolate relevance to different cases. They are intended to offer a form of rich, high-level counsel that can support evaluator or evaluation team efforts to embrace and make decisions within the unfamiliar and undetermined. Christie (2011) recently reinforced this point. “In the context of educational evaluation, the development of empirically grounded theory . . . will assist in advancing more systematic and contextually relevant evaluation practice, as well as lead to the development of contingency theories that specify the conditions under which particular evaluation practices are optimal” (p. 1). The principles introduced here stem from the systematic analysis of the stories of evaluators who have worked in a wide variety of settings and thoughtfully reflected on the nature and implications of their collaborative approaches.

This same thoughtful reflection from both evaluation theorists and practitioners is what buttresses the Program Evaluation Standards, 3rd Edition (Yarborough, Shulha, Hopson & Caruthers, 2010). These standards already serve as an effective primary resource for evaluator thinking and practice generally. It was the particular complexity inherent in evaluator/stakeholder interdependence and the influence this interdependence has on more conventional evaluation practices, however, that led us to envision a complementary resource such as principles.

The timing for this project seemed optimal given that collaborative approaches are now embedded in an abundance of well-defined evaluation models. These include, community based (e.g., Mark & Shotland, 1985), fourth generation (e.g., Lincoln, 1989), participatory (e.g., Cousins & Earl, 1992; Cousins & Whitmore, 1998; King, 2007), transformative (e.g., Cousins & Whitmore, 1998; Mertens, 2009), deliberative democratic (House & Howe, 2000), empowerment (Fetterman, 2001; Fetterman & Wandersman, 2005), and collaborative (O’Sullivan, 2004; Rodriguez-Campos, 2005) to name a few. Each of these approaches provides evaluators with a strategic plan for their work and assistance with decisions of practice.

While acknowledging the importance of these unique approaches, like Dahler Larson (2009) we argue that, “ . . . the same evaluation model or approach would probably work differently depending on the political, strategic, cultural and organizational conditions under which it is applied, and evaluators would be intelligent if they kept their evaluation practices flexible and adaptive to the varying contexts” (p. 312). We have taken the stance that collaborative approaches are likely most powerful when they remain responsive to the purpose and context for the requested evaluation and to the needs and capacities of stakeholders. We contend that fidelity to the processes and strategies associated with a single model have the potential to obfuscate the need for evaluators to be
continuously adaptive to the social, historical, ecological, and cultural complexities of the evaluation context (Cousins, Whitmore & Shulha, 2013). For this reason we maintain that collaborative approaches to evaluation stand to benefit greatly from being informed by principles.

In answering concerns expressed by Fetterman, Rodriguez-Campos, Wandersman and O’Sullivan (2014) we clarified this position. We confirmed our support for all those making efforts to advance the theory and practices associated with collaborative approaches, and expressed little doubt that access to well-defined methods, procedures and processes serves a significant need for many practitioners. Evidence for this is seen in the degree of interest in workshops and resources that provide training and professional development associated with these approaches. However, our case remains firm. During evaluations – especially those in which processes and products are grounded in close human interaction – there are likely to be continuous contextual disturbances. If the selected approach does not address the specifics of how to proceed in the light of such disturbances, evaluators may either feel compelled to stay the course or feel lost at sea as to how to proceed (Cousins, Whitmore, & Shulha, 2014). Either way, the outcome is bound to be less than desirable.

General evidence-based principles for collaborative approaches to evaluation are intended to yield guidance rather than direction. Such guidance will be particularly informative when the need for a collaborative approach is being deliberated, when evaluators are not familiar with or wedded to a specific approach, and where the complexity of the program and its environment requires flexibility and adaptation of the collaborative inquiry (Cousins, Whitmore, & Shulha, 2014).

We agree with Brown (2013) that the hallmark of professional evaluator decision making is “the enactment of choice among alternative courses of action made in response to perceived changes in circumstances and conducted in a context of ambiguity” (p. 2). Consequently, the value of these principles will rest in their capacity to illuminate complexity rather than resolve it, to inform decisions rather than prescribe them.

And so, we introduce a set of empirically grounded principles that individually, and as a set, show promise in guiding evaluation practice in contexts where the meaning of evaluation is jointly constructed by evaluators and stakeholders. We make no claims that this set as it stands today is either exhaustive or enduring. It is, however, a product of the collective wisdom of evaluators who both embrace collaborative approaches in their practice and took the time and interest to help inform how these practices could be understood. Our hope is that as collaborative approaches are both used and refined, empirical and scholarly work will seek to test the veracity of these principles. In encouraging such work, we acknowledge that the principles themselves are likely to evolve.

**Methodology**

Data collection encompassed four phases represented in Figure 1. The initial phases involved on the development and pilot of the data collection instrument as well as focus groups to test the desirability of developing a set of overarching principles for collaborative inquiry. The third phase was an online survey that drew on the expertise of practicing evaluators to identify dimensions, factors or characteristics that enhance or impede success in collaborative approaches. The fourth and final phase was a validation exercise of the draft principles emerging from phase 3 findings. We used mixed methods to gather data but the principles presented below were mainly derived from qualitative responses to queries about project success or lack of it.
Phase 1: Pilot of Data Collection Instrument
- Co-principal investigators generated narratives around successful and unsuccessful collaborative approaches to evaluation from their experience.
- Narratives were analyzed for factors enhancing or impeding success.

Phase 2a: AEAThink Tanks
- Group discussions about the merit of a set of principles.
  - Fall 2011: Ideas from working groups and individual participants collected.
  - Fall 2012: Individual thinking about successful and unsuccessful projects followed by small group discussions about reasons for success, or lack thereof.
  - Responses subsequently analyzed for factors enhancing or impeding success.

Phase 2b: Pilot Test of Draft Survey (2013 spring-summer)
- Design and administration of the draft online survey to willing Think Tank participants requiring responses about successful and unsuccessful collaborative approaches to evaluation.
- Revisions made to the draft online survey.

Phase 3: Online Survey of Evaluation Practitioners
- Invitations sent to evaluation practitioners and theorists in the AEA, CES and IDEAS.
- N=320 usable responses received. Qualitative data examined for factors contributing to success or lack of success.
- Identification of major themes and sub-themes led to the development of a draft set of principles.

Phase 4: Validation of Draft Set of Principles
- Draft set of principles circulated among interested Phase 3 participants for review and feedback via an online survey.
- Content analysis conducted of survey responses for each of the principles as well as the set of principles as a whole.
- Revisions made to the draft set of principles.
- The research team was able to confirm the proposed overarching principles, dimensions and contributing factors.

Figure 2: Sequential phases of the study from 2011 to 2014
Phase 1: Pilot of Data Collection Instrument

The principal investigators (Cousins, Whitmore, & Shulha) piloted the methodology in the first phase. Each independently reflected on (i) a successful collaborative evaluation in which they were involved, and (ii) a collaborative evaluation that they considered to be less than successful. While some consideration was given to describing the project and its context, the main focus was on the reasons for success or absence of it.

Data from this pilot phase were content analyzed in terms of their yield of dimensions, and factors or characteristics that enhanced or impeded success in collaborative approaches to evaluation. The principal feature of the methodology was the identification and comparison of factors supporting success against those leading to a lack of success in collaborative approaches to evaluation. While many factors are no doubt unique to one condition or the other, it seemed likely that some factors directly differentiate successful and less than successful projects; the presence of a given factor may contribute to success in one case but its absence may lead to lack of success in a separate project. We were of the view that such factors may be critical in determining project success and therefore potentially fertile grounds for considering principle development. Eventually a survey tool was developed based in part on the findings of this phase.

Phase 2a: American Evaluation Association (AEA) Think Tank Sessions

In November 2011 and October 2012, we coordinated think tank sessions at the annual conference of the AEA, both sponsored by the Topical Interest Group: Collaborative, Participatory and Empowerment Evaluation. The purposes of the think tank sessions were to consider the desirability of developing a set of overarching principles for collaborative inquiry to evaluation and to further pilot the methodology described above. Approximately 25 members attended each of the respective sessions. The 2011 Think Tank session consisted of discussions about collaborative approaches to evaluation within working groups, with results recorded on flipcharts; some individual participants also submitted feedback sheets. The 2012 Think Tank session followed a similar format but discussions focussed mainly on factors that contribute to the success or lack of success of evaluation projects that use collaborative approaches of inquiry. Responses from both Think Tank sessions were subsequently content analyzed for factors enhancing or impeding success of collaborative approaches to evaluation and contributed to the refinement of the data collection instrument. The majority of AEA members involved in the 2011 and 2012 Think Tank sessions voluntarily provided contact information and agreed to participate in subsequent phases of the research project.

Phase 2b: Pilot Test of Survey with Think Tank Participants

In Phase 2b, we invited volunteering participants from both the 2011 and 2012 Think Tank sessions to pilot test the survey instrument using online questionnaire survey tool. The survey required some quantitative but mainly qualitative responses about (i) successful and (ii) less than successful evaluation projects using collaborative approaches to inquiry. Quantitative ratings corresponded to collaborative project purposes and processes and were consistent with the framework developed by (Cousins & Whitmore, 1998). Our main focus, however, was on qualitative open-ended responses that explained why the project was “successful” or “far less successful than you had hoped”. Based on feedback and a low response rate to the pilot exercise, we further revised and shortened the length of the survey questionnaire in preparation for phase 3 of the research.
Phase 3: On-line Survey of Evaluation Practitioners

The draft principles were mainly derived from the responses to the open-ended questions concerning successful and less than successful collaborative approaches to evaluation (Table 1). The order of the question sets (successful, less than successful) was randomized to counterbalance for possible response set bias.

<table>
<thead>
<tr>
<th>Focus</th>
<th>Survey Questions</th>
</tr>
</thead>
</table>
| **Successful Collaborative approach to Evaluation** | 1. What are the top 3 reasons why this collaborative approach to evaluation was highly successful?  
2. Provide more detail about the project (e.g., purpose, context, and other reasons). |
| **Unsuccessful approach to Collaborative Evaluation** | 1. What were the top 3 reasons why this collaborative approach to evaluation was far less successful than you had hoped?  
2. Provide more details about the project (e.g., purpose, context, and other reasons). |

We requested permission for access to mailing lists from three professional evaluation societies (i) the AEA, (ii) the Canadian Evaluation Society (CES) and (iii) the International Development Evaluation Association (IDEAS). We were granted permission by AEA and IDEAS but were required to use a mass broadcast approach for CES members. Regardless of approach, we provided a reminder about two weeks after the initial invitation. Participation was strictly voluntary and anonymous. However, in the final section of the questionnaire we asked if respondents would be interested in participating in Phase 4 of the study and if so, to specify their email address for future contact.

We ultimately received N=320 useable responses from practicing evaluators. It should be noted that, as is the case for many online surveys, it is not possible for us to calculate a response rate because not all participants actively engage with collaborative approaches to evaluation and many society members do not even practice evaluation (e.g., trainers, program and organization decision makers).

All of the respondents to our questionnaire necessarily practice or have practiced evaluation regardless of whether they primarily self-identify as an evaluator. By far and away the majority of the responses came from AEA members (N=297 or 93%). In the online survey we asked AEA respondents to identify the Topical Interest Group (TIG) to which they are primarily identify. Among the leading TIGs in terms of representation were Collaborative, Participatory and Empowerment (9%), Independent Consulting (8%), PreK-12 Educational Evaluation (7%), Health Evaluation (6%) and Evaluation Use (5%).

Table 2 shows that most of the sample (72%) self-identified as evaluation practitioners, while the remaining respondents were from a variety of other roles such as teachers/trainers (6.9%), researchers/theorists (10.0%), commissioners or overseers of evaluation (3.4%), and other (6.9%). The majority of the sample (54.2%) had over 11 years of experience working as an evaluator.
The open-ended survey data were downloaded into Word files and then imported into NVivo 10 for analysis. We began the analysis with independent line-by-line coding, looking for emerging themes that contributed to successful and less than successful collaborative inquiry, but also differentiated successful from less than successful. Two data analysts using an iterative process performed the analysis of data. This involved reading the data multiple times, coding it, developing initial themes and holding regular meetings amongst co-researchers. Our objective was to obtain as high an agreement as possible among co-researchers coding the data. We approached this challenge using the following protocol: (i) co-researchers regularly selected a random segment of data and coded it individually, then subsequently met to compare and discuss interpretations, resolve discrepancies (e.g., codes added or deleted, codes merged, definitions clarified), and ultimately to develop a coding structure and a codebook; (ii) once we were satisfied with our coding structure and agreement, we conducted a quality check which included the two data analysts and a co-principal investigator independently coding a randomly selected segment of data (97 responses). We developed and applied a scoring system/rubric (presented in Table 3) to assess inter-coder agreement.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Three co-researchers coded a response using exactly the same code(s).</td>
<td>Excellent</td>
</tr>
<tr>
<td>2</td>
<td>Three co-researchers coded a response using the same code, but also used additional different codes.</td>
<td>Very Good</td>
</tr>
<tr>
<td>1</td>
<td>Two of co-researchers coded the response using the same code, but the third coder used a different code.</td>
<td>Good</td>
</tr>
<tr>
<td>0</td>
<td>None of the co-researchers used the same codes.</td>
<td>Unsatisfactory</td>
</tr>
</tbody>
</table>

Table 3: Coding Comparison Rubric

Our inter-coder comparison indicated that 19 responses (20.4%) scored 3 points (excellent), 41 responses (44.1%) scored 2 points (very good), 23 responses (22.5%) scored 1 (good) and 14 responses (15%) scored 0 (unsatisfactory). Most of our coding had good to excellent inter-coder agreement. Upon closer examination, we discovered that the sections that had unsatisfactory inter-coder reliability were mainly a result of the use of codes that had similar meanings (e.g., interpersonal relationships and evaluator/stakeholder relationships). Once we identified the overlapping codes, we discussed their meaning, merged some codes and finalized our codebook. It was this codebook that was subsequently used by the two data analysts in coding the remaining data.
Our next step was to conduct a coding matrix query that involved a cross-tabulation of the coding intersection between the data sets of successful and less than successful collaborative approaches to evaluation. This query confirmed that the majority of the overarching themes emerged equally in both sets of data (i.e., presence contributed to the success or absence of success in using a collaborative approach to evaluation). The team of five researchers then met face-to-face for an examination the results. An in-depth review of the results from this first level of analysis yielded a set of eight overarching categories as well as themes and sub-themes.

From there, the researchers worked in dyads to draft a set of principles of collaborative approaches to evaluation. Each principle included descriptive dimensions (themes) and contributing factors (sub-themes) that helped to illustrate a rich meaning for each respective principle. Included with the set of principles were data (verbatim quotations) that typified the language and experiences underpinning associated contributing factors. The majority of the contributing factors include verbatim quotations from both successful and less than successful responses; however, a few contributing factors emerged exclusively from the less than successful data (i.e., sponsor expectations include engagement in a collaborative approach) which explains why some verbatim quotations only come from the data of less than successful collaborative projects.

**Phase 4 Validation of Draft Set of Principles**

A summary paper that explicated the draft principles for collaborative inquiry was circulated among all Phase 3 participants who indicated a willingness to participate in the validation of the draft set of principles. In addition to receiving the summary paper, participants were invited to respond to an online survey to provide feedback on the draft principles. Participants were asked to comment on the principles as well as on the suggested direction for ongoing inquiry and field testing.

Of the 320 persons responding to Phase 3, 279 (87%) volunteered to participate in Phase 4. Data were collected over a three week period within which one reminder was sent. Ultimately, we received responses from 65 (23.3%) respondents but only 57 actually completed the survey. A 58th participant provided extensive feedback in the form of an email, since the online data collection platform had been closed before he could respond. Respondents were predominantly experienced with slightly more than half having worked as an evaluator for more than 15 years. Most self-identified as evaluation practitioners (43 or 75.4%), while 6 were teachers/trainers, 3 were researchers/theorists, 1 was an evaluation commissioner and 4 indicated ‘other’. The median number of evaluations on which respondents participated as either principal evaluator or team member was 25 (min 2, max 400).

In-depth content analysis was conducted for all the responses for each of the principles as well as the set of principles as a whole. Our research team reviewed the analysis, and systematically refined each of the principles based on the feedback, if justified by the data. Based on this validation phase, we identified some minor challenges to the clarity and contents of the principles that warranted changes, but overall we were able to confirm the proposed overarching principles, dimensions and contributing factors. Table 4 shows the results of quantitative ratings of the draft phase 3 principles. It is quite evident that participants both supported each of the draft principles as important guidance to practice and that they were of the view that intended meaning of the principles was clear.
## Table 4: Participant Ratings of Draft Phase 3 Principles

<table>
<thead>
<tr>
<th>Draft Phase 3 Principle</th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Clarifying impetus for Collaboration</td>
<td>6.59</td>
<td>0.68</td>
<td>56</td>
<td>5.88</td>
<td>1.21</td>
<td>56</td>
</tr>
<tr>
<td>2: Understanding Program Characteristics and Contexts</td>
<td>6.46</td>
<td>0.76</td>
<td>56</td>
<td>5.95</td>
<td>1.20</td>
<td>56</td>
</tr>
<tr>
<td>3: Assessing Evaluation Valuing</td>
<td>6.33</td>
<td>0.95</td>
<td>54</td>
<td>5.57</td>
<td>1.79</td>
<td>54</td>
</tr>
<tr>
<td>4: Ensuring Adequate Resources</td>
<td>6.43</td>
<td>1.13</td>
<td>54</td>
<td>6.39</td>
<td>1.09</td>
<td>54</td>
</tr>
<tr>
<td>5: Practicing Participatory Processes</td>
<td>6.52</td>
<td>0.75</td>
<td>54</td>
<td>6.33</td>
<td>1.23</td>
<td>54</td>
</tr>
<tr>
<td>6: Growing Collaborative Relationships</td>
<td>6.47</td>
<td>0.89</td>
<td>53</td>
<td>6.09</td>
<td>1.21</td>
<td>53</td>
</tr>
<tr>
<td>7: Assuring Evaluation Quality</td>
<td>6.57</td>
<td>0.80</td>
<td>53</td>
<td>6.40</td>
<td>1.15</td>
<td>53</td>
</tr>
<tr>
<td>8: Maximizing Positive Evaluation Consequences</td>
<td>6.34</td>
<td>1.07</td>
<td>53</td>
<td>6.26</td>
<td>1.27</td>
<td>53</td>
</tr>
</tbody>
</table>

### Principles to Guide Collaborative Approaches to Evaluation

The principles described below are not presented in any order of importance. As illustrated by Figure 1, the principles are conceptualized as a set of interdependent considerations that are relevant when using collaborative approaches in evaluation. The importance of interconnectedness became evident in our analysis, where participants’ quotations about their collaborative experiences were occasionally used to support the development of more than one principle. In the ensuing presentation we draw from participants’ verbatim quotations to support each of the principles. Refer to Appendix A for the final set of principles with examples of verbatim quotations that supported the development of the principle.

Two considerations are pivotal to the use of the principles. First, we contend that the list of principles is not a menu from which evaluators ought to choose in undertaking collaborative work. Adherence to individual principles is a matter of degree, as opposed to a ‘whether-or-not’ proposition. In short, we
see each principle as being essential to evaluation practice in a collaborative context; the extent to which any given principle is important will depend entirely on contextual conditions, circumstances and complexities. For an elaborate discussion of the essentiality of principles, see Patton (2015b).

A second consideration is that the principles should not be prioritized a priori; we make no claim as to their temporal order. To reiterate, we have reasoned that a decision about which principle to emphasize and when, is likely to be contingent on the purpose of the evaluation, the stage of the evaluation, the context in which the collaborative approach is being implemented, and the emergence of complexities as the evaluation unfolds.

With this in mind, we now turn to a brief explanation of each principle. Note that contributing factors associated with each principle appear in the diagrammatic expert from Figure 1 used to identify each principle.

**Principle Descriptions**

As evaluators and stakeholders move towards undertaking a collaborative approach, significant attention should be paid to what this means in practice and why such an approach is desirable. While on the face of it, the need for these understandings appears self-evident, our respondents reminded us that policies mandating collaborative evaluation are not unusual, and the charge to collaborate certainly does not guarantee collaborative practices in action. Not engaging in a frank, joint discussion with stakeholders around a request or proposal for a collaborative approach can be costly. Learning too late that “the grant application articulated a collaborative approach that the managers did not support,” can railroad not only any hope for a meaningful collaboration but the chance for a meaningful evaluation as well.

The clarification process suggested by this principle is the task of making understandings around the purpose of the evaluation, the information and process needs embedded in the purpose, and the expectations of stakeholders around the collaboration explicit and transparent for all collaborators. As one evaluator told us, this is one way to have “all stakeholders understand and be deeply committed to the project.”

The presence of a “global misunderstanding of the evaluation purposes for both the evaluation team and the main stakeholders” can leave “staff, [feeling] not part of the evaluation committee, and [as though] it is not part of their job to participate.” When the purpose is agreed upon, “stakeholders add additional perspectives of what is feasible to implement, and what they are willing to commit to implementation.”

Some purposes appear more conducive to the use of collaborative approaches than others. When the collaborative approach worked well, we were often told that the program improvement,
opportunities for individual and organizational learning and organizational capacity building were high on the list of purposes. We also heard that,

The program managers greatly valued the potential for the evaluation to help them continuously improve the program.

*The stakeholders were very interested in determining how their program worked, how to improve the program, and how to demonstrate outcomes.*

*Everyone had a common goal of program improvement.*

*The staff was committed to learning how to conduct an evaluation.*

In contrast, when accountability and legitimizing purposes were the focus of the evaluation, it was more problematic for the collaboration. In one context, “government renewal of the program was at stake in the evaluation and this made the program participants quite defensive.” Another respondent lamented, “data were used to justify the program rather than improve it.”

Clarifying the expectations around collaborative approaches to evaluation requires learning about the extent to which stakeholders will welcome engagement and be prepared to work at fostering the approach. In one instance success in the collaborative approach was attributed to “stakeholders [who] leveraged their network for the benefit of the design, data collection, and validation of findings.” In contrast, one funder “used the evaluation to force collaboration . . . among stakeholders.” Sensing reluctance and coercion at the program site early in the process and then taking time to understand and temper these dispositions may be one way to avoid intentional or unintentional sabotage of the evaluation by stakeholders down the road. An example illustrated this clearly: “The program developers and implementers wanted to do their own evaluation. . . . When we were hired they participated in but changed everything - this was not positive participation.”

Not only is it important to establish the meaning of the collaborative approach early, there are benefits to reinforcing this meaning over time. Doing so appears to provide a touchstone as the evaluation unfolds. As one respondent put it, “…we both knew we were committed to collaboration, which didn’t make it easy, it just meant we both kept that goal salient when things got tricky”. Nurturing a collaborative approach can begin with an agreement about how information gets communicated among those actively involved. “The organization ensured the evaluator [he/she] could speak independently with all stakeholder groups.” In comparison, another evaluator described how the collaboration was thwarted because she/he did not or could not “fully tap into motives of various stakeholders to voluntarily offer information and participate in the evaluation.”

Evaluators citing successful collaborative approaches have probed, documented, and shared among stakeholders the information and process needs underpinning their work together. “The clients knew what they wanted to understand, but didn’t know how to find out the answers, so I was able to work with them to sort through their ideas and set priorities.” Sometimes these needs and priorities are discovered through formal evaluation activities. “Our stakeholders are always involved in identifying the evaluation questions and this [was] successful to make sure we [were] collecting information that [would] be useful to them.” Not agreeing on the needs to be addressed can have serious consequences in how evaluators and stakeholders perceive the quality of the collaboration and ultimately the evaluation. “Since everything was a priority and everything got measured, nothing was done well.” Another evaluator lamented, “we couldn’t acknowledge or work through differences or prioritize anything.”

Taking what is learned about information and process needs and, through joint effort translating these needs into the evaluation design also enables evaluators and stakeholders to examine
assumptions about how collaboration might meet these needs. “People are more likely to buy into this [collaborative] process when they have a hand in designing it.” As much as evaluators may be open to joint design, they may encounter stakeholders who are less than enthusiastic. “Not all stakeholders shared a sense of urgency or expected value added by conducting an evaluation.” Left unattended, this may be an indication of a root problem requiring attention. “Program stakeholders did not want the evaluation to occur. [We learned] it had been requested by outside management.” Engaging stakeholders early and often about their motivation for working collaboratively is no guarantee of a successful collaboration. Evaluators have told us, however, that processing information and evidence about the proposed purpose for the evaluation, identifying the information and process needs underpinning the call for evaluation, and clarifying client/stakeholder expectations around how the evaluation will unfold can help both evaluators and stakeholders to articulate the assumptions and needs that the call for collaboration is intended to address.

Successful collaborative approach to evaluation, we were told, relies on the quality of the relationships that evaluators and stakeholders are able to develop and sustain. When reflecting on these relationships one evaluator described how the evaluation “. . . was conducted in a highly cooperative, and collaborative organizational context, with abundant positive peer/professional relations and a wholesome, trusting, organizational climate.” Valuing and utilizing the contributions of stakeholders appears to be central to a context of mutual respect. One evaluator talked about how critical the understandings of stakeholders were to their team; “the stakeholders knew much more about the context of the evaluation than we could absorb in the short time we were on the job.” At the same time evaluators recognized that their ability to establish credibility with new stakeholders may be, in part, a function of previous processes: “The stakeholders valued our relevant prior experience on very similar projects. They knew we had a reputation for doing good work on this type of project. They were very open and eager to work with us.” For evaluators, the danger in assuming that reputation can automatically foster a strong inter-professional relationship is in not remembering the attention, skills and the effort that were central to developing this reputation.

Trust between evaluators and their stakeholders is not cemented by a contract. Trust requires purposeful effort on all sides and is more effective when the effort is transparent, as one evaluator shared “evaluators need to demonstrate that they want to, and are listening.” It was encouraging to hear that when respect and trust are in place, it helped clients/stakeholder to “. . . be honest with the evaluators about their strengths and more importantly weaknesses.” In at least one case, trust and respect also facilitated patience with the collaborative process: “The genuine effort to be collaborative was recognized by partners even if it wasn’t always perfectly implemented as a collaborative process by the evaluator.” In such a context evaluators must avoid “too many unspoken assumptions.” Instead, as one evaluator reported, when energy was invested to “work out differences directly with each other . . . learning was less threatening.”

“We may be the evaluation experts, but they are subject matter [program content] experts and both are very important.”
It was clear from evaluator stories that the quality of an emerging relationship does not rest solely in the hands of the evaluator. Rather than proceed in the face of resistance, however, it may be wise to revisit the initial motivation for a collaborative approach, and the evaluation itself. Some evaluators lamented the fact that despite their best efforts, “stakeholders did not act as partners and had very little trust in the evaluation process” and “program developers weren't willing to engage with evaluators.” And finally, “from my standpoint, no matter how accessible I tried to be, no matter how many pies and cookies I baked for team meetings, staff still saw me as an outsider whose main job was to check up on them.”

Huberman (1999) documented the influence of sustained interactivity on research utilization. Respect and trust also appears to benefit from this same form of structured and sustained interactivity. For example, one respondent shared that “close and constant contact was instrumental to real-time communication and relationship building.” We heard about the importance of having “clearly defined and communicated expectations, roles and responsibilities.” With these in place it is no doubt easier to value and use the “frequent feedback from program managers and local evaluators during all stages of the evaluation.”

“Cultural competence is a stance taken toward culture, not a discrete status or simple mastery of particular knowledge and skills” (AEA, 2011, p. 1). It is not surprising to learn that relationships are reinforced when the collaborative approach is monitored for its capacity to acknowledge, respect, and honour diversity. Evaluators who felt their collaborative approach to evaluation had been successful described different approaches to adopting this stance. One purposefully worked through “a values and needs identification process that was inclusive and participatory.” Another put together a “multicultural team of evaluators all of whom were skilled evaluators in their own right as well as bringing different cultural lenses.” A culturally competent stance can only work in favour of a collaborative approach. Such a stance was reported to be central in working with “an innovative program in a strongly cultural space. Without collaboration [the evaluation] would have been largely meaningless, as cultural stakeholders wanted their say.”

Becoming culturally competent in evaluation requires evaluators to “maintain a high degree of self-awareness and self-examination to better understand how their own backgrounds and other life experiences serve as assets or limitations in the conduct of an evaluation” (AEA, 2011, p. 1). If we adopt Gray’s 1989 notion that collaboration is “a process through which parties who see different aspects of a problem [or issue] can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible” (p.5) then the ability to build and maintain respectful and valued relationships can be viewed as an expression of cultural competence in evaluators.

“The evaluator helped project leaders articulate their program objectives and learning outcomes for participants.”

All evaluations are encouraged to “document programs and their contexts with appropriate detail and scope for the evaluation purposes” (Yarbrough et al., 2010, p. 185). When the approach to evaluation is collaborative, engaging stakeholders in documenting the goals, objectives, and intended
implementation of the program was reported to help the evaluation itself run more smoothly. Typically stakeholders are included in the program description process: “The involvement of stakeholders provided a more accurate definition of the terms, problems, and population needs [and] culture.” A less conventional way to go about this task according to one respondent involved, “conducting interviews with program participants [which] allowed funder stakeholders to understand how the program worked.” Another cautioned against this approach because “the program was in transition and difficult to find a consistent thread/voice among program participants.” No matter how the description process takes place, focusing on a mutual understanding of what is being evaluated can reduce the likelihood of stakeholders moving forward in the evaluation with “unrealistic expectations about the program outcomes/design.”

Practicing evaluators also told us there is value in moving beyond the elements of a logic model as a guide for describing and discussing the program. A more shared understanding of both the organizational context within which the program is operating and the organization’s capacity for engaging in a collaborative approach to evaluate may be equally important. Successful collaborations were connected to working with “a program manager who was intent on making sure that her program was successful, constantly improving and had the documentation to prove it” and with “supervisors [who] supported program developers, implementers and front-line staff to have time to work on evaluation.” In one practitioner’s experience, it was unclear whether an evaluation was ever conducted, but the option for collaboration for this evaluator certainly vanished when disturbing evidence about the program’s organization led to this judgment: “Institutional racism [was] very strong and embedded throughout the bureaucratic system. [It was] perpetuated in research and evaluation.”

It was not uncommon for the evaluators in our study to begin their collaborative work with stakeholders feeling confident in the capacity of the organization to embrace the process. The problem was that over the life of the evaluation, this capacity seemed to diminish or disappear. We were told about “a mid-project change in administration [that] decreased political support for the project, [and] the motivation for stakeholders to participate” and, how “significant organizational turnover occurred at the dissemination and use phase, so new leadership wanted to follow a new vision rendering the work irrelevant.” Whether or not evaluators have the wherewithal to mitigate the erosion of commitment to evaluations grounded in collaboration will depend both on the skills of the evaluator and emerging conditions within the organization. In any case, continuous monitoring of the organization can alert evaluators to conditions that may erode the willingness or ability of stakeholders to engage.

“Staff was involved at all levels from design decisions, to data collection, to findings presentations.”

What does it mean for stakeholders to be “involved” in a collaborative approach to evaluation? One way to answer this question is to refer to the significant work that has been done to name, define,
document, differentiate, and compartmentalize specific forms of evaluations that promote a collaborative approach (Fetterman, Rodriguez-Campos, Wandersman, & O’Sullivan, 2014). Alternatively, if we adopt Gray’s (1998) notion of collaboration as cited above, involvement in a collaborative approach to evaluation can be operationalized more responsibly. Being “involved” in a collaborative approach then becomes defined by the context specific decisions and processes that evaluators and stakeholders make, in order to use their differing visions, in search of optimal ways to address identified information and process needs. Regardless of whether the collaborative approach is predetermined, emergent or some combination of both, decisions will need to be made about the optimal form of participation; specifically, who will participate in the collaboration, how those identified will participate, and who will have control over decision making during the various phases of their joint effort (Cousins & Whitmore, 1998; Cousins & Chouinard, 2012). References to these essential facets of participation infused the experiences that evaluators shared with us.

Evaluators in this study typically attributed the degree of success they experienced in using a collaborative approach to the way one or more of these three dimensions played out. Where they varied was in where along these dimensions they chose to operate. For example, one evaluator described how

…participants were close to - and ultimately owned - the data. They helped design the tools, collect the data, analyze the data, interpret the data, and presented findings. It wasn’t just buy-in to the process and outcome – it was implementing the process themselves (not being led through) and generating (not being given and asked for their thoughts about) and owning the outcomes.

This description suggests that, at least in that context, the success of the collaborative approach was attributable to in-depth participation by stakeholders who were also invested with a significant amount of decision-making authority.

Many evaluators reported on the importance of identifying and considering a variety of stakeholders, especially when these individuals or groups “otherwise might not have been involved.” Doing so, made it possible to “prevent a few stakeholders and other decision makers from remaining silent partners.” One evaluator was proud that for the first time in a specific context, “the project had a more in-depth process to hear beneficiaries’ voices.” We heard, however, that the actual challenge might not be in identifying the diversity of stakeholders most appropriate for inclusion in the collaborative approach, but in negotiating their participation. “Decision makers did not want those best equipped to contribute . . . to participate as anything more than sources of data - following marching orders.”

The optimal depth of participation of stakeholders will be different depending on the purpose or form of collaborative approach. There was evidence, however, that in evaluations considered to be successful, those who were identified as being central to the collaboration were also engaged in shaping the approach. In one instance, “stakeholders and the evaluator participated in conceptualizing the project, before it was even funded.” The evaluation appears to benefit when these same stakeholders are engaged in meaning making at critical points along the way: “Preparing reports involved stakeholders so that multiple ways of sharing evaluation information was done. These methods included briefs, interim reports at various milestones, presentations, and final report with an executive brief.” We were also told of a successful collaborative approach where, “stakeholders interpreted the findings, generated and implemented recommendations for change, and presented the findings to their colleagues.” While there is no rule for how deeply stakeholders should participate in the evaluation, the worst-case scenario appears to be when there are “unclear expectations for participation in evaluation activities.”

Decision making within a collaborative approach may just be the most difficult participatory process to manage. One evaluator did attribute the success of the evaluation directly to the “evaluator [being]
open to sharing the control of the evaluation, particularly as it related to instrument choice and
development, data collection methods, and interpretation of data.” Many more, however, attributed
their less than successful experience in implementing a collaborative approach to evaluation to the
complications that arose around control over decision making. In these examples it is interesting to
note that the power issues were not always between the evaluator and stakeholders.

All stakeholders are not created equal. Some have greater influence over others and do believe their
voice should carry greater weight in articulating the evaluation findings.

The advisory group did not have any real influence to change the evaluation design or the survey
assessment tools. My sense is that they perceived their lack of influence and the group did not meet
as a group beyond one or two meetings.

The steering committee was not representative, and unduly influenced or even forced the whole
process.

Program developers wanted too much involvement and did not have the skills or experience to
help with the evaluation; they co-opted the evaluation and then used it to report what they wanted
to report, rather than reality.

Funders ultimately were poor collaborators. [They] dominated . . . power dynamics [were]
dysfunctional.

A vulnerable group [was able] to drive the evaluation. Because they represented the majority
participants, at any meeting, they were able to control data interpretation and meaning making.

Instead of an evaluator leading the stakeholders through a well thought out process, the program
team was somewhat in charge and the evaluator was at their mercy.

The ways in which these evaluators had envisioned decision making being distributed within their
collaborative approaches are not clear. What is clear are their frustrations at not being able to either
create the conditions necessary to put their plans into place or to respond to the conditions that
arose. Being insistent in learning about how program and policy decisions within an organization are
made and about the presence and authority of a program’s informal leadership are two keys that may
provide early insight into the challenges that may arise around control over decision making.

“**The program manager was able to devote 25% of paid work time towards the evaluation . . . It was not something extra.**”

The full cost required to support collaborative approaches to evaluation, including budget, time, and
skilled personnel can remain abstract to both evaluators and client/stakeholders, sometimes until well
into the process. Typically this set of resources works together to shape the feasibility of the
collaborative approach. A change in one can dramatically influence the others.

Only the evaluator headlining this principle made reference to a purposeful redistribution of funds in
support of the time and effort required of the stakeholders engaged in the collaboration. It was more
common to hear from evaluators that there were “insufficient funds to meet the expectations for the evaluation”. If the collaboration is identified as ‘part of the job’ for those who will be heavily involved, then asking what will be removed from the list of their responsibilities during the evaluation may be a way to revisit the purpose of and expectations around the collaborative approach.

Even when everyone is confident that the evaluation has been appropriately funded, fiscal conditions are known to change. Evaluators told us about how a “budget crisis led to an evaluation cutback.” Such changes are often beyond the control of those collaborating. But evaluation funds can also be vulnerable in the face of existing or emerging organizational priorities. “The funded initiative was the result of available, external funding. Although the management had agreed the [evaluation] work was important, really they just wanted the money.”

It is not unusual for collaborative approaches to require more time to implement than conventional approaches. This appears to be especially true when there are process and capacity building goals. Stakeholders may need to negotiate the timing of their engagement or be coached in the skills defining their participation. One evaluator attributed collaboration and evaluation success to “taking the time not to rush the process and [providing] ample opportunity for partners to provide feedback.” The fact that the evaluator was willing to take the time necessary for meaningful collaboration “was recognized and appreciated by partners.” Being conscious of externally imposed time constraints and the implications on time for growing the collaboration and nurturing the appropriate inquiry skills is critical. One evaluator told us about how success was constrained by, “limits [that] existed around the school year.” Another lamented, “the yearly evaluation [was] based on a funding cycle [that was] too short to capture more qualitative changes.”

In an evaluation built on a collaborative approach the most important resource may be the people who are working together. Many evaluators talked about having difficulty motivating stakeholders to stay engaged, and it was not because the evaluation was going badly. Instead the collaboration suffered from emerging conditions within the evaluation context. For example,

- [There was] sporadic participation by stakeholders . . . because of schedule conflicts and work demands.
- The program [was] coming to an end and staff [were] looking for other job opportunities.
- [There was] staff turnover due to job instability within the organization. [This] significantly limited the ability to engage more stakeholders.

Evaluators seemed to have sympathy for collaborators who faced such conditions. For example, “[our] client was stretched too thin and was operating in a highly political environment, which affected her ability to spend sufficient time with the evaluation team.” They were less compassionate when they felt the motivation of those collaborating was drained by clients using the evaluation as a ‘catch all’ for their information needs. “The organization . . . kept adding tasks to the evaluation plan which burdened evaluators and other stakeholders and did not have buy-in.”

Whether grounded in a specific and well-defined collaborative approach or evolving from an examination of evaluation purposes, needs, expectations, and context, a collaborative approach works best when the personnel collaborating can, as a team, bring to the table an appropriate combination of facilitation and inquiry skills. Predicting what these skills might be can help to establish realistic expectations for roles and responsibilities. “The evaluator was not an expert in the program content area and absolutely needed stakeholders to provide clarity about how the data would be used and what the boundary conditions were for asking questions of intended beneficiaries.”
Assessing the extent to which the skills critical to the approach are present is one step. One evaluator said, “the program managers did not have adequate data management skill” another noted that, “the evaluation team did not have breadth of experience to deal with complexity of program.” Deciding what to do about the need for more expertise in a supposedly collaborative approach is more problematic, in which one fallback position is “more reliance on the evaluator for their skill sets and expertise.”

“Due to lack of experience on my part as a young evaluator I did not engage the client in troubleshooting problems as they arose as much as I should have and they spiralled out of control, resulting in inconsistent data collection across sites, and problems with the quality of the data collected.

Evaluations will progress towards important findings and useful outcomes as long as the design driving the evaluation remains credible. Evaluation design decisions are typically the most appropriate and powerful at the moment they are agreed upon. Our participants reminded us that as the evaluation unfolds, however, it is not unusual for the evaluation context (i.e., people, programs, organizational context, information needs, and process needs) to change. The continued implementation of early design decisions, therefore, may actually be problematic in moving the evaluation forward. Frank conversations at these times allow for modifications that can keep the evaluation on track. For example, “[the] climate in which the evaluation was being conducted was conducive for talking about disagreements openly and honestly and coming to a resolution that everyone was okay with moving forward.”

Evaluators talked about the need to remain “flexible [and] very open to [making] adjustment/changes in the evaluation design and implementation.” Staying attuned to the consequences of early design decisions, “on an ongoing basis”, can alert the evaluator and client/stakeholders about the need to re-envision the path forward. Acknowledging and sometimes confronting each other with any accelerating lack of fit between the intended evaluation design and the capacity of the collaboration to implement this design can be productive and evaluation saving. As one respondent reported “the evaluation design changed and adapted to include negotiations around stakeholder needs and requests.” A preliminary assessment of the organization’s stability and evaluation history may help to avoid a situation such as this one: “The evaluation design, specified in a grant proposal, [became] awkward and burdensome, and it could not be changed after the grant was awarded.”

Evaluators in this study identified issues around data collection as the most common threats to evaluation progress and technical quality. Evaluators are typically well trained in how to create logical and methodologically sound evaluation designs – ones that will lead to reliable/dependable data and justified/trustworthy findings. Our participants revealed that issues around accuracy (Yarbrough et al., 2011) are often no different in evaluations using collaborative approaches than they are in non-collaborative designs. For example, “sadly, data requested were not accurate because stakeholders were “hiding” duplication and overlap of services for their clients” and “evaluation data, by the time it reached those who could use it best, was so massaged by the client . . . (it had to look just so) that it was watered down and meaningless.”
It is apparent, however, that there are some unique pitfalls when purposefully engaging client/stakeholders in data decisions. Assuming that client/stakeholders are appreciative of the implications of data quality on findings and outcomes may be the first of these. “Front line staff, who were responsible for collecting the data, did not understand the importance of getting it collected accurately.” Also problematic are situations where collaborators see response rates as the gold standard for assessing the quality of their work in developing or administering data collection tools. For example, “the staff changed instrumentation during the collection phase but did not inform the evaluator about the need to change the instrument.” This situation may have been triggered by a legitimate concern for those acting as data sources and how they were interacting with the instruments. Another evaluator commented, “instruments used in evaluation were interpreted like a test to be passed.” Not all stakeholders will be engaged in shaping elements of the evaluation even when a collaborative approach is guiding decision making.

It is critical to remember that individuals and groups whose experiences, ideas and feelings are central to answering the evaluation questions are also stakeholders. Guiding collaborators on how to help those offering data to see the importance of their role and the logic behind the data collection tools may be necessary. Doing so will improve the likelihood that data collection instruments are working as intended and may even build evaluation capacity in the collaborators themselves.

In an ideal world, the resources to build the skills of those involved are integrated into the budget. For example, “the stakeholders provided the personnel to train and collect our data systematically across all sites and followed up on any data collection problems that we identified.” A more common story is of stakeholders striking out with minimal or no preparation. As a consequence we were told that, “some of the changes to the evaluation made by stakeholders resulted in questions that were duplicative or difficult to interpret” and “measurable indicators were poorly chosen and did not answer critical questions about program benefits.” In the absence of formal training of stakeholders clarifying the assumptions that stakeholders have concerning the value of data collection instruments and the processes of data collection are recommended, such attention may reduce the amount of monitoring necessary during data collection itself and go a long way in preserving the integrity of the evaluation.

Archibald (2013) defined evaluative thinking as, “an attitude of inquisitiveness and a belief in the value of evidence, that involves skills such as identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection, and perspective taking and making informed decisions in preparation for action” (November, 2013, par. 5). Preskill and Torres (1999) described how evaluative thinking requires “members [to] come together to engage in the learning processes of: (1) Dialogue, (2) Reflection, (3) Asking Questions, and (4) Identifying and Clarifying Values, Beliefs, Assumptions, and Knowledge” (p. 45). When evaluative thinking is done collaboratively, it is intended to make evaluation processes and findings “more meaningful to stakeholders, more useful for decision
makers and more effective within an organization” (Torres et. al, 2000, p. 28). The strong relationship between inquiry and learning already established as part of formal evaluative thinking was verified by participants in this project.

It was common to see the successes attributed to both the collaborations and their evaluations connected to a commitment to inquiry. For example, “the organization put evaluation [as] the top priority. [They were] willing to spend time on it“ and “the evaluation had the buy-in of agency management and staff at all levels.” Specific reference was often made to the disposition of stakeholders towards evidence-informed decision making:

*The program managers have used evaluative data for program improvement and to communicate program successes to stakeholders.*

*The program administrator was committed to collecting data and using information for program improvement, accountability and future funding purposes.*

*They] believed in the value of using a rigorous evaluation design and data collection/analysis procedures.*

It became apparent that in some contexts the evaluation was adopted as a process to nurture a culture of inquiry for the organization. Evaluators in our study were encouraged when “the agency leadership supported and encouraged all levels of staff to participate.” In one case, evaluation became “a frequent topic during staff meetings.” In a second, “inclusion of the program recipients in the data collection was deemed necessary by the program sponsors and stakeholders.” In a third, the evaluator being encouraged to engage “key champions among stakeholders for evaluation.”

How does an evaluator help stakeholders to keep the inquiry focused on learning? This was the challenge faced by at least one evaluator in our study who described “low interest in learning about project successes and challenges. . . . [It is] easier to implement [a program] when you [think you] already know it, than to modify and change according to the findings, conclusions, recommendations.” From what participants in this study said, the key may be to suppress any early inclinations to assure collaborators of the current merit, worth and significance of their program. In more successful contexts, significant energy seems to have been spent helping collaborators first becoming invested in the learning process and being prepared for the unexpected. For example,

*Because of the stakeholder commitment, results were used as an opportunity to learn and grow.*

*Program designers and implementers were open to feedback and recommendations based on data from the evaluation.*

*Stakeholders were willing to accept negative or contrary results without killing the messenger.*

In the absence of a strong orientation of stakeholders and their organizations towards evaluative thinking (inquiry and learning), evaluators may need to adopt, at least initially, an educative role in their collaborative approach (Shulha & Cousins, 1997; Johnson et al., 2009). Doing so may help overcome some of the barriers that evaluators in this study traced to a fundamental distrust of evaluation. “People unfamiliar with evaluation, were defensive and in denial of any need for improvement.” In one instance this was attributed to a lack of experience. “[There was] a lack of a basic understanding about evaluation activities amongst clients.” Once again, anxiety around evaluation is not unique to those using collaborative approaches, but when present and left unattended, stakeholders will have few reasons to trust either the evaluation or the collaboration.

The educative role may be equally important when it is apparent that stakeholders have misunderstandings about how evaluation might work for them. In one context, an evaluator reported that “the culture of the group receiving the evaluation was not one that fostered data use. They saw
the evaluation as being for ‘someone else’ or as a ‘proof of concept’ to justify expenditures.” In another, “line staff were not invested in the evaluation and had no interest in learning about evaluation because they felt it did not apply to their lives. Once the grant was over, they felt they’d never see evaluation again.” Evaluators may need to draw liberally on the strategies suggested by Archibald (2013) and Preskill and Torres (2000) if they are to help stakeholders move beyond initial obstructing assumptions. In the absence of strong evidence that stakeholders and their organizations are invested in evaluative thinking, the experiences of evaluators who participated in this study suggest that it would be wise to cultivate it.

Kirkhart (2000) was central in helping evaluators expand their vision of the consequences of evaluation for persons and systems. Our data were not rich enough to examine how all three of Kirkhart’s (2000) suggested dimensions (source, intention, and time) may have been molded by the collaborative approach. They did reveal, however, that evaluating the outcomes dimensions, as stimulated both by the evaluation results and processes, were shaped by collaboration. The influence of collaborative approaches on outcomes, were easily characterized using a framework proposed by (Cousins & Whitmore, 1998). Practical outcomes at the organizational level influence program, policy and structural decision-making. For stakeholders, practical outcomes are seen through a change in disposition toward the program or evaluation, and the development of program skills including systematic evaluative inquiry. Transformative outcomes are connected primarily to a change in the way organizations and individuals view the construction of knowledge and the distribution and use of power and control. Such changes are labeled transformative only when they promote the type of social change that enhances independence and democratic capacities of ordinary people. The strength of this framework is its recognized capacity to capture outcomes that can surface in participatory contexts and to stimulate thinking about how to proceed with collaborative approaches when these outcomes are a priority (King, 2007).

In this study, collaboration influenced practical outcomes directly through the way findings were derived. By working together with the data, “stakeholders were better able to consider multiple reasons for results.” One evaluator identified the need for buy-in from those not normally targeted for participation. Expanding the stakeholder group can be an appropriate adaptation as the evaluation unfolds. “Representatives from among the program’s intended beneficiaries were actively involved in data collection and in reporting results; the evaluation [findings] had greater success in getting a serious hearing when program decisions were made.” When client/stakeholders were not engaged in the production of findings, it was more common to hear of a “miscommunication or misunderstanding of what the final product (i.e., the report) could tell us.” One evaluator lamented, “we tried to get them engaged in interpreting results but to no avail. The management team had no time to give to . . . meaning making.”

“There was a great deal of process learning: stakeholders were given a chance to reflect on their program by being involved in creating the logic model and identifying evaluation questions. They said often that this was just as, or even more helpful than the final report summarizing findings.”
One positive but possibly complicating consequence of collaborating around data is that stakeholders may form conclusions around findings as the evaluation proceeds. Helping stakeholders suspend judgments until the data are in can be challenging. Acting on partial answers to evaluation questions, however, can be perilous for practical outcomes, as the following comments suggest.

*The agency did not wait long enough for results, decided that early intervention had succeeded, adopted a version system wide, and saw costs balloon because the intervention did not work.*

*…the results created problems for senior management. They revealed that major barriers to applying evidence-based practices were their internal systems.*

When evaluation findings are not proving to be as encouraging as anticipated, decision makers may use the time between data and findings to prepare a response that is not evidence informed. One evaluator told us, the “results were not positive, and in their disappointment, program management dismissed [their] credibility.”

Possibly the most disappointing experience for evaluators using a collaborative approach is to discover that, despite best efforts, there was “no real intent to do anything with the results. Once [the evaluation] was complete, they could get back to business as usual.” For those confronting this kind of frustration, documentation throughout the evaluation on how the collaborative approaches were conceptualized and implemented can at least enable a productive internal meta-evaluation of assumptions and decisions (Yarbrough et al., 2010). This form of continuous professional learning typifies evaluators who grow, rather than repeat, their practice. Despite these pitfalls, collaborative approaches appear to enhance opportunities for practical outcomes, as seen in this example: “Participatory data interpretation process led to insights about program improvement that were immediately adopted and implemented.”

Transformative outcomes have arisen in contexts where evaluative inquiry has been used to deliberatively engage stakeholders in the social construction of knowledge, and when these emerging understandings were then used both to propel the evaluation forward and to re-shape the program and the organization (Harnar, 2012; Monkman, Miles & Easton, 2007; Flores, 2007). Our participants confirmed this stance even though they did not explicitly identify transformative outcomes as a goal for their work. At the organizational level, our participants reported that “working collaboratively deepened the sense of community among the stakeholders,” and that “stakeholders became more empathetic to intended beneficiaries of results due to understanding the complex nature of problems.”

Transformational outcomes were implied more when the facilitating evaluator appeared to be skillful in promoting inquiry and have expertise in human and social dynamics. Being prepared to work towards transformational outcomes almost certainly means being prepared to work in contexts where there are differences and even conflict. For example, “meetings were held with stakeholders with different viewpoints and interests, and when the program started there was some tension among them. Meetings and co-creation of evaluation promoted meaningful discussion, understanding and helped to reduce tension.” In another context, “the collaborative approach was highly successful because it supported relationship building between the two communities. The processes revealed that the two communities had not had formal relationships for decades and this needed to be addressed.”

Evaluators in this project confirmed a relationship between the use of collaborative approaches and the appearance of practical and transformational outcomes. We realize, however, that these types of outcomes are not totally independent (in that what is practical can also be transformative and vice versa). When these evaluators chose a collaborative approach to evaluation, it was evidence of an
intentional effort to generate outcomes that would have immediate and positive consequences for stakeholders and their organizations. Disappointment in their ability to consistently accomplish this goal suggests that evaluators would be wise to negotiate with stakeholders (i) the range of outcomes possible given the scope of the evaluation, (ii) which outcomes are most worthy of purposeful attention, and (iii) how joint effort might best facilitate these outcomes.

Moving the Principles Forward

The principles as derived and described above reflect the understandings that current evaluation practitioners bring to the complexity of implementing collaborative approaches to evaluation. We introduce these principles, not as advocates for or owners of them. Rather, we encourage colleagues – academics, practitioners, commissioners of and stakeholders in evaluation – to take these principles forward and make them “the subject of continuous analysis and renewal through dialogue and systematic inquiry” (Cousins, Whitmore, & Shulha, 2013, p. 18). To be relevant, these principles must show potential for adding value to our collective efforts. To be powerful, the nuances of interdependence need to be made even more explicit. The true test of these principles will be the extent to which they are able to strengthen collaborations, and enhance evaluator working knowledge.

We also challenge readers to keep in mind the boundaries of the knowledge that has shaped them. For any set of principles to be truly appropriate for collaborative work, stakeholders need to be included in the conversation. It is a point not lost on the authors. The principles could, no doubt, benefit greatly from having wide variety of stakeholders and evaluation teams reflecting systematically on the complexities of interdependence (personal communication, Jill Chouinard, November 2014). Such efforts could shed important light on how these guiding principles might best be refreshed. Other questions that remain unaddressed or only partly addressed include,

• How specifically do evaluators define success in collaborative approaches to evaluation?
• How do these principles align with specific collaborative approaches of evaluation? Do some principles more than others complement these approaches?
• To what extent are these principles responsive to contextual complexities?
• Do the principles resonate well in varying cultural contexts?
• Are these principles useful to novice evaluators, or are they best reserved for evaluators with experience?

Our sense is that the principles, used as a set to guide and reflect on collaborative practice hold strong potential for enhancing the success of such evaluations and we encourage ongoing, well documented field trials to confirm this hunch. But there are potential benefits of the principles. Within the realm of evaluation education and training, one can easily imagine the development of instructional modules around each principle. Might they not also serve as an interesting basis for organizational evaluation policy reform? Regardless, it is our conviction that the principles require solid test-driving opportunities and they should be revisited and perhaps re-engineered sometime not too far down the road.

In closing, we want to underscore the contributions of the significant group of evaluators who helped to shape this research and the hundreds more who were willing to contribute their experiences. Our effort to unearth these principles was sustained by the commitment of colleagues to the project. We regard this as the beginning of a dynamic and evolving conversation and look forward to the dialogue.
References


## Appendix A: Building Principles to Guide Approaches to Evaluation

**Principle: Clarify Motivation for Collaboration**

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
</table>
| **Evaluation purpose** | **Expectations for the evaluation are clear for evaluators and stakeholders.** | - S-Our stakeholders are always involved in identifying the evaluation questions.  
- S-All stakeholders understand and are deeply committed to the project.  
- S-We had common goals and were committed to a process of program development.  
- Stakeholders add additional perspectives of what is feasible to implement, and what they are willing to commit to implementation.  
- U-Global misunderstanding of evaluation purposes for both the evaluation team and the main stakeholders.  
- U-If purpose is purely accountability and process is dictated from funder, collaborative approaches are more difficult to implement.  
- U-The staff . . . not part of the evaluation committee.  
  . . felt it was not their job to participate.  |
|                      | **Evaluation purposes center around program improvement, learning and capacity building.** | - S- The program managers greatly valued the potential for the evaluation to help them continuously improve the program.  
- S The stakeholders were very interested in determining how their program worked, how to improve the program, and how to demonstrate outcomes.  
- S-Capacity building- the staff were committed to learning HOW to conduct an evaluation.  
- S- Everyone had a common goal of program improvement.  
- U-[Evaluation only for accountability, not for learning.  
- U- Government renewal of the program was at stake in the evaluation and this made the program participants quite defensive.  
- U - Data were used to justify the program rather than improve it.  |
<table>
<thead>
<tr>
<th>Information and process needs</th>
<th>The information and process needs are probed, documented, prioritized and agreed upon.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• S-The clients knew what they wanted to understand, but didn't know how to find out the answers, so I was able to work with them to sort through their ideas and set priorities.</td>
</tr>
<tr>
<td></td>
<td>• S- Our stakeholders are always involved in identifying the evaluation questions and this is successful to make sure we are collecting information that will be useful to them.</td>
</tr>
<tr>
<td></td>
<td>• U- Since everything was a priority and everything got measured, nothing was done well...there was no compromise for the sake of quality.</td>
</tr>
<tr>
<td></td>
<td>• U- We couldn't acknowledge or work through differences or prioritize anything.</td>
</tr>
<tr>
<td>Evaluation design reflects joint effort to meet identified information and process needs.</td>
<td>• S-People are more likely to buy in to this process when they have a hand in designing it.</td>
</tr>
<tr>
<td></td>
<td>• S- All funders stakeholders were involved in the design, commissioning and management of the evaluation and expected to be involved in the evaluation process.</td>
</tr>
<tr>
<td></td>
<td>• S - Tons of stakeholder involvement helping that design.</td>
</tr>
<tr>
<td></td>
<td>• U- Program stakeholders did not want evaluation to occur (had been requested by outside management).</td>
</tr>
<tr>
<td></td>
<td>• U- Buy in- not all stakeholders shared sense of urgency or expected value added by conducting an evaluation.</td>
</tr>
<tr>
<td>Evaluator and stakeholder expectations</td>
<td>Client and stakeholders welcome engagement in a collaborative approach.</td>
</tr>
<tr>
<td></td>
<td>• S- Stakeholders leveraged their network for the benefit of the design, data collection, and validation of findings.</td>
</tr>
<tr>
<td></td>
<td>• S- Recognition of key program leaders that collaboration is an essential way of working.</td>
</tr>
<tr>
<td></td>
<td>• U- Grant application articulated a collaborative approach that the managers did not support.</td>
</tr>
<tr>
<td></td>
<td>• U- Funder used the evaluation to “force” collaboration not only among stakeholders in this project, but also demanded collaboration with four other collaborative evaluations.</td>
</tr>
<tr>
<td></td>
<td>• U- The program developers and implementers wanted to do their own evaluation and not have my group as an outside evaluator; when we were hired they participated in but changed everything - this was not positive participation.</td>
</tr>
</tbody>
</table>
Intent and meaning of the collaborative approach is established early and reinforced over time.

- S- the sponsor of the evaluation wrote a RFP asking specifically for a collaborative approach and I described in detail what that meant to me.
- S- When I was selected to do the evaluation we both knew we were committed to collaboration (which didn’t make it easy, it just meant we both kept that goal salient when things got tricky).
- S- Open nature of communication with stakeholders at all stages of the evaluation.
- Open nature of communication with stakeholders at all stages of the evaluation
- The organization ensured the evaluator could speak independently with all stakeholder groups.
- U- We didn’t fully tap into motives of various stakeholders to voluntarily offer information and participate in the evaluation for the good of the whole.
- U- Evaluator separated from funder by client who did not let information flow.
- U- Participation by key stakeholders waned as the process advanced.

**Principle: Foster Meaningful Relationships**

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
</table>
| Respect, Trust       | Evaluators and stakeholders identify, value and respect the contributions they each will make to the evaluation. | • S- It was conducted in a highly cooperative, collaborative organization context with abundant positive peer/professional relations and a wholesome, trusting organizational climate. 
  • S-The evaluator and stakeholders contributed in unique ways. 
  • S-We may be the evaluation experts, but they are subject matter experts and both are very important. 
  • S- The stakeholders knew much more about the context of the evaluation than we could absorb in the short time we were on the job. 
  • S-The stakeholders valued our relevant prior experience on very similar projects. They knew we had a reputation at NSF for doing good work on this type of project. They were very open and eager to work with us. 
  • S-Good working relationship between the evaluation team, program administrators, and program staff. 
  • S-Program manager and implementers respected the evaluator’s expertise and were open to suggestions. |
<table>
<thead>
<tr>
<th>Structured and Sustained Interactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Efforts to establish trust are purposeful.</strong></td>
</tr>
<tr>
<td>• S- High level of trust among collaborators.</td>
</tr>
<tr>
<td>• S-Necessary to gain stakeholder trust. They need to be honest with the evaluators about their strengths and more importantly weaknesses and lessons learned.</td>
</tr>
<tr>
<td>• S-The evaluators need to demonstrate that they want to and are listening.</td>
</tr>
<tr>
<td>• U-It seemed as though the decision-makers did not value or trust the evaluation team.</td>
</tr>
<tr>
<td>• U-Stakeholder did not act as partners and had very little trust in the evaluation process.</td>
</tr>
<tr>
<td>• U-Program developers weren't willing to engage with evaluators.</td>
</tr>
<tr>
<td>• U-From my standpoint, no matter how accessible I tried to be, no matter how many pies and cookies I baked for team meetings, staff still saw me as an outsider whose main job was to check up on them.</td>
</tr>
</tbody>
</table>

| Evaluators and stakeholders are transparent in their efforts to be collaborative. |
| • S-Genuine effort to be collaborative was recognized by partners even if it wasn't always perfectly implemented as a collaborative process by the evaluator. |
| • S-Because two communities were collaborating, it allowed them to share all sorts of information about how they were approaching the work they were mutually engaged in. |
| • S-Worked out differences directly with each other. Learning was less threatening. |
| • S-Open nature of communication with stakeholders at all stages of the evaluation. |
| • U-Each stakeholder had its own agenda regarding renewal of the grant and each department was in “competition” |

| Structures required to facilitate interactions, both within and across roles, are established and reinforced. |
| • S-Because two communities were collaborating, it allowed them to share all sorts of information about how they were approaching the work they were mutually engaged in. |
| • S-Worked out differences directly with each other. Learning was less threatening. |
| • S-Open nature of communication with stakeholders at all stages of the evaluation. |
| • U-Each stakeholder had its own agenda regarding renewal of the grant and each department was in “competition” with each other. |

| Methods for sustained interaction are identified and agreed upon. |
| • S-Clearly defined and communicated expectations, roles and responsibilities. |
| • S-Clear roles and responsibilities, so input and expectations are managed properly. |
| • S-Close and constant contact was instrumental to real-time communication and relationship building. |
| • S-The evaluation team spent 3/4 of its time at the site. |
| • S-Frequent feedback from program managers and local evaluators during all stages of the evaluation. |
### Cultural Competency

- S-Continuous feedback to stakeholders (primarily the funder) about evaluation results, to help inform continuous program improvement.
- U-Too many unspoken assumptions.
- U-There seemed to be miscommunication among staff about purposes (e.g. there was good staff to evaluator communication but not staff to staff).

Collaborative approach is monitored for its capacity to acknowledge, respect and honour diversity.

- S-An innovative program in strongly cultural space - without collaboration would have been largely meaningless, as cultural stakeholders wanted their say.
- S-Multicultural team of evaluators all of whom were skilled evaluators in their own right as well as bringing different cultural lenses.
- S-Culturally appropriate evaluation - the instruments developed with the help of the stakeholders.
- S-Went through a "values and needs identification" process that was inclusive and participatory.
- U-The evaluator who was hired to conduct the evaluation was not culturally sensitive.
- U-Language barrier.
- U-Institutional racism is very strong and embedded throughout the bureaucratic system and is perpetuated in research and evaluation.

---

**Principle: Develop a Shared Understanding of the Program**

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
</table>
| **Program Logic**    | The program is described in detail. | • Project goals, objectives, implementation, and corresponding evaluation went smoothly.  
  • S-Evaluator helped project leaders articulate their program objectives and learning outcomes for participants at the start of the program.  
  • S-Conducting interviews with program participants allowed funder stakeholders to understand how the program worked.  
  • S-The involvement of stakeholders provided more accurate definition of the terms, problems, and population needs/culture for more accurate reporting of results.  
  • U-Program was in transition and difficult to find a consistent thread/voice among program participants.  
  • U-Unrealistic expectations about program outcomes/design. |
| **Organizational Context** | Program managers explicit in their | • S-Excellent program manager who was intent on making sure that her program was successful, constantly improving and had the documentation to prove it. |
| Committing to the program and its evaluation. | S-Supervisors supported program developers, implementers and front-line staff to have time to work on evaluation. |
| Organizational capacity for collaborative engagement is monitored. | U-Significant organizational turnover occurred at the dissemination and use phase, so new leadership wanted to follow a new vision rendering the work irrelevant. |
| | U-A mid-project change in administration decreased political support for the project, decreasing the motivation for stakeholders to participate. |

### Principle: Promote Appropriate Participatory Processes

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diversity of stakeholders</strong></td>
<td>A variety of stakeholders are identified and considered for participation.</td>
<td>S- Engaged a variety of stakeholders that otherwise might not have been involved.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-Every voice was heard - no fear in sharing information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S- Prevented (a few) stakeholders and other decision makers from remaining silent partners,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-First time that the project had a more in-depth process to hear beneficiaries' voices.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U-Decision-makers did not want those best equipped to contribute even to the design to participate as anything more than sources of data following marching orders.</td>
</tr>
<tr>
<td><strong>Depth of participation</strong></td>
<td>Stakeholders central to the evaluation are engaged in shaping the collaborative approach.</td>
<td>S-The inclusion of the program staff as part of the collaborative team during each phase of the evaluation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-Stakeholders and evaluator participated in conceptualizing the project, before it was even funded.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-Consistent engagement throughout the process.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The project was a partnership and in order to provide optimal services, all partners had to be included in most, if not all, steps of the evaluation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S- The evaluation design was constantly reviewed, revised and discussed during regular meetings with the client.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-Staff was involved/informed at all levels from design decisions to data collection to findings presentations. Evaluation was a frequent topic during staff meetings.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U-unclear expectations for participation in evaluation activities.</td>
</tr>
<tr>
<td></td>
<td>Stakeholders central to the collaborative approach are engaged in meaning making</td>
<td>S-Preparing reports involved stakeholders so that multiple ways of sharing evaluation information was done. These methods included briefs, interim reports at various milestones, presentations, and final report with an executive brief.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S-The involvement of stakeholders provided more</td>
</tr>
<tr>
<td>Control of decision making</td>
<td>at critical points in the evaluation.</td>
<td>accurate ... reporting of results.</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Evaluator(s) and stakeholders negotiate appropriate influence in determining decision points and decision-making processes.</td>
<td>• S-Evaluator was open to sharing the control of the evaluation, particularly as it related to instrument choice and development, data collection methods, and interpretation of data.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-stakeholders did not have enough of a voice in decision making.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-Although it was loosely designed to be collaborative, the advisory group did not have any real influence to change the evaluation design or the survey assessment tools. My sense is that they perceived their lack of influence and the group did not meet as a group beyond one or two meetings.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U- instead of an evaluator leading the stakeholders through a well thought out process, the program team was somewhat in charge and the evaluator was somewhat at their mercy.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-Program developers wanted too much involvement and did not have the skills or experience to help with the evaluation; they co-opted the evaluation and then used it to report what they wanted to report, rather than reality.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U- It enabled a vulnerable group to drive the evaluation. Because they represented the majority participants, at any meeting, they were able to control data interpretation and meaning making.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-Certain stakeholders controlled the process. Other stakeholders were not able to bring their thoughts/concerns to the project.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-All stakeholders are not created equal. Some have greater influence over others and do believe their voice should carry greater weight in articulating the evaluation findings.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-The steering committee was not representative, and unduly influenced or even forced the whole process.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U- Funders ultimately were poor collaborators; dominated; power dynamics dysfunctional.</td>
<td></td>
</tr>
<tr>
<td>Contributing Factors</td>
<td>Themes</td>
<td>Data</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>------</td>
</tr>
</tbody>
</table>
| **Budget**           | The evaluation budget remains realistic given the investment in a collaborative approach. | - S-Taking the time not to rush the process and provide ample opportunity for partners to provide feedback and have input was recognized and appreciated by partners.  
- U- yearly evaluation based on funding cycle too short to capture more qualitative changes.  
- U- This was a complex evaluation involving multiple methods over a limited time period (24 months).  
- U- Limits existed around the school year. |
| **Time**             | Time frame for the evaluation supports joint effort. | - S-Program manager was able to devote 25% of paid work time towards the evaluation...it was not something "extra" that was put on her plate.  
- S- sufficient resources to support participation across stakeholder groups.  
- U-Budget crisis led to evaluation cutback.  
- U- Insufficient funds to meet the expectations for the evaluation.  
- U- The funded initiative was the result of available, external funding. Although the management had agreed the work was important really they just wanted the money. |
| **Personnel**        | Expectations around stakeholder participation are realistic. | - U- Sporadic participation by stakeholders -- rarely more than 50% participated because of schedule conflict and work demands.  
- U- The project/program is coming to an end and staff are looking for other job opportunities. Having staff around to facilitate the collaboration of various stakeholders is a challenge.  
- U- Staff turnover due to job instability within the social service organization significantly limited the ability to engage more stakeholders and serve more clients.  
- U- Bad timing: staff and political changes.  
- U- Organization ... kept adding tasks to the evaluation plan which burdened evaluators and other stakeholders and did not have buy-in.  
- U- Client was stretched too thin and was operating in a |
highly political environment, which affected her ability to spend sufficient time with the evaluation team.

- **U-**Ability of stakeholders to meet at the same time (the stakeholders were spread out and had limited availability).

<table>
<thead>
<tr>
<th>Evaluator and stakeholder expertise required to support the collaborative approach is accurately assessed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>**S-**Appreciation of various sets of skills and expertise (stakeholders as experts in some areas and evaluators as experts in others).</td>
</tr>
<tr>
<td>**U-**The program managers did not have adequate data management skills.</td>
</tr>
<tr>
<td>**U-**Evaluation team did not have breadth of experience to deal with complexity of program.</td>
</tr>
<tr>
<td>**U-**More reliance on the evaluator for their skill sets and expertise in proposing an initial framework, instruments, analysis plan, etc. and then providing input.</td>
</tr>
<tr>
<td>**S-**Evaluator was not an &quot;expert&quot; in the program content area and absolutely needed stakeholders to provide clarity about how the data would be used and what the boundary conditions were for asking questions of intended beneficiaries.</td>
</tr>
</tbody>
</table>

### Monitor Evaluation Progress and Quality

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaluation Design</strong></td>
<td><strong>Data</strong></td>
</tr>
<tr>
<td>Design flexible and responsive to emergent contextual conditions.</td>
<td>• **S-**Flexible; very open to make adjustment/changes in the evaluation design and implementation.</td>
</tr>
<tr>
<td></td>
<td>• **S-**The evaluation design changed and adapted to include (negotiations) around stakeholder needs/requests.</td>
</tr>
<tr>
<td></td>
<td>• **S-**Implementation data were provided on an on-going basis.</td>
</tr>
<tr>
<td></td>
<td>• **S-**Climate in which the evaluation was being conducted was conducive for talking about disagreements openly and honestly and coming to resolution that everyone was okay with moving forward.</td>
</tr>
<tr>
<td></td>
<td>• **U-**The evaluation design, specified in a grant proposal, was awkward and burdensome, and it could not be changed after the grant was awarded.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adaptations are assessed for their influence on evaluation quality.</th>
<th><strong>Data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• **U-**Evaluation data, by the time it reached those who could use it best, was so massaged (by the client system— it had to look just so...) that it was watered down and meaningless.</td>
<td></td>
</tr>
<tr>
<td>• **U-**Staff changed instrumentation during collection phase but did not inform evaluator about the need to change the</td>
<td></td>
</tr>
</tbody>
</table>
Data Collection

The consequences of engaging stakeholders as decision makers in data collection are monitored.

- U- Some of the changes to the evaluation made by stakeholders resulted in questions that were duplicative or difficult to interpret.
- U- Due to lack of experience on my part as a young evaluator I did not engage the client in troubleshooting problems as they arose as much as I should have and they spiralled out of control, resulting in inconsistent data collection across sites, and problems with the quality of the data collected.
- S- Excellent data and planned collection from the start.
- S- Very valid and reliable instruments.
- S- The stakeholders provided the personnel to train and collect our data systematically across all sites and followed up on any data collection problems that we identified.
- U- Inability to test data collection tools with population.
- U- Instruments used in evaluation were interpreted like a test to be passed rather than a set of standards to aspire to.
- U- Not sufficient methods used to capture the information needed.
- U- Measurable indicators were poorly chosen and did not answer critical questions about program benefits.
- U- Front line staff, who were responsible for collecting the data, did not understand the importance of getting it collected accurately.
- U- Capacity of different grantees to collect and report data was variable.
- U- Sadly, data requested were not accurate because stakeholders were "hiding" duplication and overlap of services for the clients.

Principle: Promote Evaluative Thinking

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inquiry Orientation</td>
<td>The value of inquiry is made explicit by program leaders.</td>
<td>- S- The organization put evaluation on the top priority. Willing to spend time on it. - S- Agency leadership supported and encouraged all levels of staff to participate. - S- Evaluation was a frequent topic during staff meetings. - S- The evaluation had the buy-in of agency management and staff at all levels. - S- Key champions among stakeholders for evaluation, particularly within funder organization. - S- Inclusion of the program recipients in the data collection was deemed necessary by the program sponsors and...</td>
</tr>
<tr>
<td>Focus on Learning</td>
<td>Program community is open to using feedback, including negative or unexpected evaluation findings.</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• S-The organization was committed to the evaluation and very open about its strengths and weaknesses in doing the work.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• S-Stakeholders were willing to accept negative or contrary results without &quot;killing the messenger.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• S-Because of the stakeholder commitment results were used as an opportunity to learn and grow.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• S-Program designers and implementers were open to feedback and recommendations based on data from the evaluation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-Line staff were not invested in the evaluation and had no interest in learning about evaluation because they felt it did not apply to their lives. Once the grant was over, they felt they'd never see evaluation again.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-People unfamiliar with evaluation, defensive and in denial of any need for improvement.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• U-Low interest in learning about project successes and challenges. Next project already started with the same</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evidence-informed decision-making is held in high regard.</th>
<th>• S-The stakeholders believed in the value of using a rigorous evaluation design and data collection/analysis procedures.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• S-Leadership committed to use results of study.</td>
</tr>
<tr>
<td></td>
<td>• S-Program Managers have used the evaluative data for program improvement and to communicate program successes to stakeholders.</td>
</tr>
<tr>
<td></td>
<td>• S-Program administrator was committed to collecting data and using information for program improvement, accountability and future funding purposes.</td>
</tr>
<tr>
<td></td>
<td>• S-Participants were close to - and ultimately owned - the data. They helped design the tools, collect the data, analyze the data, interpret the data, and presented findings. It wasn't just buy-in to the process and outcome -- it was implementing the process themselves (not being led through) and generating and owning the (not being given and asked for their thoughts about) outcomes.</td>
</tr>
<tr>
<td></td>
<td>• U-The culture of the group receiving the evaluation was not one that fostered data use. They saw the evaluation as being for &quot;someone else&quot; or as a &quot;proof of concept&quot; to justify expenditures.</td>
</tr>
<tr>
<td></td>
<td>• U-Funder did not emphasize data-based improvement or evaluation.</td>
</tr>
</tbody>
</table>
Principle: Follow Through to Realize Use

<table>
<thead>
<tr>
<th>Contributing Factors</th>
<th>Themes</th>
<th>Data</th>
</tr>
</thead>
</table>
| Practical Outcomes   | Multiple ways of interpreting findings are deliberated, when constructing the products of evaluation. | • S-Stakeholders are always involved in developing recommendations.  
• S-Stakeholders were better able to consider multiple reasons for results.  
• U-Miscommunication or misunderstanding of what the final product (i.e. the report) could tell us.  
• U-The management team had no time to give to the evaluation process or meaning making.  
• U-We tried to get them engaged in interpreting results but to no avail.  
• U-Results were not positive, and in their disappointment program management dismissed its credibility  
• U-The results created problems for senior management. They revealed that major barriers to applying evidenced based practices were their internal systems.  

Stakeholders develop capacity to integrate findings into decisions about the program and their organization. | • S- Clients had put in place processes to use the information for decision making.  
• S- S-because representatives from among the program’s intended beneficiaries were actively involved in data collection and in reporting results, the evaluation had greater success in getting serious hearing when program decisions were made.  
• Our stakeholders have an idea from the very start of how they could use the information to make program decisions.  
• S- Transfer of knowledge to front line staff regarding performance measurement and outcomes.  
• S- Made them think about how to revise program more.  
• S- Assisted in making the findings useful for program improvement and sometimes made it possible to make program adjustments based on evaluation findings.  
• S- Participatory data interpretation process led to insights about program improvement that were immediately adopted and implemented.  
• U- No real intent to do anything with the results. Once complete, they can get back to business as usual.  
• U-Agency did not wait long enough for results, decided that early intervention had succeeded, adopted a version system wide, and saw costs balloon because the |
<table>
<thead>
<tr>
<th>Transformative Outcomes</th>
<th>Individual learning and organizational capacity building are considered as products of the collaborative approach.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>intervention did not [work].</td>
</tr>
<tr>
<td></td>
<td>• U- Findings were not used extensively to improve the program.</td>
</tr>
<tr>
<td></td>
<td>• The reporting and knowledge translation part were not done properly due to a change in context and a lack of motivation from the stakeholders and participant.</td>
</tr>
<tr>
<td></td>
<td>• S- Evaluation results were viewed as learning.</td>
</tr>
<tr>
<td></td>
<td>• S- There was a great deal of process learning: stakeholders were given a chance to reflect on their program by being involved in creating the logic model and identifying evaluation questions. They said often that this was just as, or even more, helpful than the final report summarizing findings.</td>
</tr>
<tr>
<td></td>
<td>• S- Working collaboratively deepened the sense of community among the stakeholders.</td>
</tr>
<tr>
<td></td>
<td>• S- Stakeholders became more empathetic to intended beneficiaries of results due to understanding the complex nature of problems.</td>
</tr>
<tr>
<td></td>
<td>• S- Meetings were held with stakeholders with different viewpoints and interests, and when the program started there was some tension among them. Meetings and co-creation of evaluation promoted meaningful discussion, understanding and helped to reduce tension.</td>
</tr>
<tr>
<td></td>
<td>• S- The collaborative approach was highly successful because it supported relationship building between the two communities. The processes revealed that the two communities had not had formal relationships for decades and this needed to be addressed. Strategies were developed, outside of the evaluation to address this.</td>
</tr>
</tbody>
</table>