A Case Study of the Guiding Principles for Collaborative Approaches to Evaluation in a Developmental Evaluation Context

Michelle Searle
Independent Consultant

Stefan Merchant, Agnieszka Chalas, and Chi Yan Lam
Queen’s University

Abstract: Recently, Shulha, Whitmore, Cousins, Gilbert, and al Hudib (2015) proposed a set of evidence-based principles to guide collaboration. Our research undertakes a case study approach to explore these principles in a developmental evaluation context. Data were collected at two points in an 18-month period where an evaluation group collaborated with the program team from a national organization. This article explores the contributions of selected collaborative approaches to evaluation principles as they are applied in a developmental evaluation. The article concludes with a reflection on the implications for collaboration in theory and practice of developmental contexts. Also identified are the practical insights for implementing the principles in evaluation practice.

Keywords: Collaborative approaches to evaluation, developmental evaluation, education, technology


Mots clés : approches collaboratives en évaluation, évaluation développementale, technologie pédagogique

Evaluation theory has several functions; one of these is to guide practice (Mark, 2005). The integration of theory from developmental and collaborative approaches
to evaluation introduces a degree of complexity that merits further investigation when applied in practice. Patton (1994) initially described developmental evaluations (DE) as “long-term, partnering relationships with clients who are, themselves, engaged in ongoing program development” (p. 312). By its very definition, developmental evaluation is a collaborative endeavour and therefore the evaluator must be proficient in developing relationships that will lead to quality evaluations and, ultimately, quality programs.

There is no single type of evaluation that is a “collaborative evaluation.” Rather, collaboration is a central feature in a suite of theories known as collaborative approaches to evaluation (CAE). In CAE, emphasis and value are placed on joint work between stakeholders and evaluators. The overlap between DE contexts and the intentional working relationships encompassed within CAE creates opportunities for investigating applied practice. Given the vital role that collaboration must play in a DE context, it is likely that evaluators using this approach require a nuanced understanding as well as an expanded set of competencies for using their partnerships in the practice of program evaluation of evaluation theory.

One way of framing the required collaborative competencies is found in the empirically grounded principles introduced by Shulha, Whitmore, Cousins, Gilbert, and al Hudib (2015). The eight principles detail the independent and interconnected characteristics that document the goals of collaboration within the complexity of evaluation. These principles were not developed specifically for application within DE contexts. However, the contexts of evaluation work are inherently complex and dynamic; it is often necessary to integrate multiple evaluation theories to reach a desired evaluation goal (Alkin et al., 2012; Patton, 2011a).

The dynamic nature of program evaluation and theory-to-practice integration has resulted in the field of evaluation consistently calling for empirical testing of theory in practice (e.g., Scriven, 1991; Shadish, Cook, & Leviton, 1991; Worthen, 1990). Most recently, in developing these principles, Shulha et al. (2015) stated, “our hope is that as collaborative approaches are both used and refined, subsequent empirical and scholarly work will seek to test the veracity of these principles.” Accordingly, this study provides empirical examination of the principles.

We begin by positioning the authors, to make explicit the connection to Dr. Lyn Shulha and this Special Issue. Then, we provide a brief overview of the theories of DE and CAE before situating these theories in the context of this evaluation study. We outline the methods for this case study and present the results to provide empirical evidence for a selection of principles for CAE. An integrated findings and discussion section identifies ways in which applying the principles enriches understanding of evaluation theory in practice. Future implications are offered.

**POSITIONING THE MANUSCRIPT AND AUTHORS**

This article is included as part of the special issue in tribute to Dr. Lyn M. Shulha for multiple reasons. For nearly 25 years, Lyn Shulha has mentored students in assessment and evaluation at Queen’s University. As a founding member of the
Assessment and Evaluation Group, she has been instrumental in uniting scholars, students, and alumni for the purposes of contributing to the theory and practice of scholarship in program evaluation. The authors of this article, who were also evaluators on this project, are past and present students who have had the honour of learning from and working with Lyn Shulha.

The evaluation team described in this article was led by Lyn Shulha. The evaluation team members credit her mentoring, in this case and countless others, as critical to their academic and experiential learning. Her leadership within the faculty and the evaluation community has an expansive and ongoing impact by inspiring evaluators as well as organizational partners to engage in collaboration and ongoing reflective practices for the purposes of learning and innovating in the area of program evaluation.

DEVELOPMENTAL EVALUATION

DE supports the ongoing growth of a program (Patton, 2011b; Patton et al., 2015). Scholars describe it as a distinct and useful approach in evaluating programs (e.g., Lam & Shulha, 2014; Patton, McKegg, & Wehipeihana, 2016). It is suited for programs that do not have “clear, specific, and measurable goals up-front” (Patton, 1994, p. 312). DE generates information and feedback enabling developers to enact changes based on emerging information. It is well suited for situations of high complexity and early stages of innovation (Gamble, 2008). DE can commence as early as program conception with insights generated from evaluation activities facilitating purposeful changes (Lam & Shulha, 2014). This inherent flexibility means that evaluation can be ongoing and responsive while keeping the project goal in mind.

It is important to note that DE is not a substitute for program development. Rather than developing a program for a client, the evaluator is expected to work with a client to bring evaluative thinking within uncertain contexts. Specifically, the evaluator is called upon to “elucidate the innovation and adaptation processes, track their implications and results, and facilitate ongoing, real-time data-based decision-making in the development process” (Patton, 2011a, p. 4). The evaluator achieves these expectations by “facilitating processes of asking evaluative questions, applying evaluation logic, and gathering and reporting evaluative data to support [program] development in real time” (Patton, 2011a, p. 1). Hence, the relationship between the evaluator and the client or stakeholder is necessarily a collaborative one focused on innovation, inquiry, and intentional use of data to inform progress (Gamble, 2008).

COLLABORATIVE APPROACHES TO EVALUATION

King, Cousins, and Whitmore (2007) remarked that all evaluation work requires collaboration to some extent, as evaluators do not work in isolation. CAE are “growing and now include a wide range of familiar evaluation approaches and models” (Shulha et al., 2015, p. 195). These approaches and models include
participatory (e.g., Cousins & Chouinard, 2012; Cousins & Earl, 1992; Cousins & Shulha, 2006; Cousins & Whitmore, 1998), empowerment (Fetterman, 1994), and collaborative (O’Sullivan, 2004; Rodriguez-Campos, 2005; Shulha & Wilson, 2003) evaluation. These are all predicated on slightly different models, though they share the goal of engaging and involving stakeholders as a way of meeting their needs. These needs are prioritized by negotiating relationships that are sensitive to context and focused on evaluation processes that promote deeper levels of stakeholder understanding (Cousins, Whitmore, & Shulha, 2013).

Methodologically, CAE suggests that rigour is found in crafting evaluation questions collectively, joint meaning-making activities as well as thoughtful analysis, dissemination, and reflection (e.g., Cousins, 2003, 2007; Cousins, Whitmore, & Shulha, 2013; Poth & Shulha, 2008). Stakeholder involvement differs by evaluator as well as evaluation context. Ultimately, emphasizing joint ways of knowing, creative forms of understanding, and shared ways of making meaning goes beyond methodological choices. The principles for CAE also call evaluators to question their epistemological positioning (Shulha et al., 2015). We assert that CAE can be employed within a DE because of their shared vision for evaluators and stakeholders to work in partnership using dynamic and emergent processes.

The principles for CAE were crafted over a four-year multiple methods study from the collective wisdom of evaluation practitioners and scholars (Shulha et al., 2015). In total the eight principles are conceptualized as “a set of interdependent considerations” (Shulha et al., 2015, p. 198). The principles are intentionally not presented in a linear or sequential manner to emphasize the complex and dynamic nature of evaluation. The principles are as follows: “clarify motivation for collaboration, follow through to realize use, promote evaluative thinking, monitor evaluation progress and quality, monitor and respond to the resource availability, promote appropriate participatory processes, develop a shared understanding of the program, and foster meaningful relationships” (Shulha et al., 2015, p. 194). Our study investigates these principles in practice to ascertain how they shaped our understanding of collaboration in a developmental context.

EVALUATION CONTEXT

The aim of the DE project was to use blended learning to engage teachers in the knowledge work and skill development required to improve students’ inquiry and questioning skills in science. A blended learning community was developed and facilitated as a pilot project between July 2014 and July 2015. This style of professional learning is a form of innovation for in-service teacher learning that is self-paced and resource-rich, and can be highly interactive/adaptive as a project progresses (Christensen, Horn, & Staker, 2013). Teachers initially participated in a full day session in person before moving online to participate. At the end of the project, teachers gathered in person to share their learning. The pilot was funded by a grant from the provincial Ministry of Education.
This evaluation context is best described as nested; the pilot project was nested within the Ministry’s other funded projects for the public education system. The pilot project was also nested within the organization’s suite of science programs. The pilot program team was a subset of the organization’s staff, including a new project lead. The external evaluation team was nested within the university and the organization team.

Within these nested entities, there were complexities and uncertainties. For example, the pilot project was a yet-to-be-developed vision, being led by someone who was new to the organization and who was also a novice in blended learning. The project requirements were partially mandated (e.g., inclusion of French and English sites; geographical locations) by the funder, and these changed part way through the grant. Other complexities were interwoven into these entities; for example, both the evaluation team and the organization’s project team brought varying levels of professional experience and backgrounds as they worked from geographically disparate locations, none of which intersected with the field sites for data collection. Ethics provided another challenge: this was the first time the organization had navigated school districts to receive ethical permission, and the university ethics application was fraught with tensions over the evaluation-research paradigms. Lastly, many school districts were engaged in labour disruptions as approvals were being secured and later when the data was being collected. In spite of all of these complexities, there was a high level of enthusiasm and vested interest on behalf of the stakeholders, evaluators, and participants. We argue that these complex layers and the dynamic state of the project made CAE within a DE particularly salient.

METHODS

A case study approach was selected because the principles put forward by Shulha et al. (2015) arose out of the collective experiences of the 320 evaluators who responded to their survey, but generalized conclusions resulting from aggregated data are not necessarily helpful in specific circumstances. A case study allowed us to determine the relevance and utility of the principles in a specific evaluation context. The case study provides a frame for data collection and analysis (Stake, 1978, 1980, 2004, 2005; Yin, 1994, 2009). All methods complied with the ethical standards outlined by the affiliated university Tri-Council Policy Statement, the organization, and the school districts.

Participants and Data Collection

The participants were six members of the university-based DE team and seven participants from the organization. A qualitative, multiple methods approach was adopted (Stake, 1995). The primary data collection methods were interviews and a focus group. During the DE process, members of the evaluation team interviewed seven participants from the organization (see Table 1). These seven participants included all of the organizational staff working on the project. A standardized
The interview protocol was followed to ensure consistency of questions, but the interview style was conversational. Interviewers posed follow-up and clarification questions where appropriate. Furthermore, because the interviews took place during the DE (and therefore during project development), the respondents often had questions for the evaluators, leading the conversation in unanticipated directions. A timeline is provided in Table 1 to show the DE and study activities.

Once the evaluation report had been submitted and the organization had time to read and comment on its overall structure and content, a focus group was conducted with four members of the evaluation team and five members of the organization. Two researchers who played minor roles in the evaluation facilitated the focus group. Using advice from Krueger and Casey (2009), the focus group facilitators developed and led the discussion, checked for understanding, and provided summaries of findings as the focus group proceeded. Facilitators developed prompts in advance; evaluation team members had no input or knowledge of the questions in advance.

The focus group interwove multiple activities as a means of augmenting the traditional focus group structure. This allowed for the collection of rich data, while honouring the relationships and dialogic processes established between the evaluation team and the organization’s participants. There were three main activities: (a) image-elicitation, (b) a placemat activity, and (c) a question and answer session.

Image-elicitation uses images as stimuli for evaluation thinking (e.g., Dempsey & Tucker, 1994; Harper, 2002; Hurworth, 2004a, 2004b; Walker, 1993; Wang & Burris, 1997; Weber, 2008). The focus group facilitators brought a deck of 140 cards, each of which had an image on it. For each of the three initial questions 60 images were chosen at random and placed on a large table. Participants were put into pairs, and each person was invited to select an image that best represented their thinking on that particular question. Pairs then used the images in describing their ideas with one another while conversations were recorded. Then, as a larger group, participants displayed their images and shared their thinking while making connections and raising questions across the group. Many more images

Table 1. Timeline for data collection

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb. 2014</td>
<td>Project proposal is accepted by funder</td>
</tr>
<tr>
<td>Mar. 2014</td>
<td>Evaluation team hired</td>
</tr>
<tr>
<td>Jun.–Aug. 2014</td>
<td>Evaluation design negotiated and accepted</td>
</tr>
<tr>
<td>Aug.–Oct. 2014</td>
<td>Project designed and online presence developed</td>
</tr>
<tr>
<td>Nov. 2014</td>
<td>Project launched with school districts</td>
</tr>
<tr>
<td>Apr. 2015</td>
<td>Interviews conducted with 7 project staff</td>
</tr>
<tr>
<td>May 2015</td>
<td>Evaluation data collection finishes</td>
</tr>
<tr>
<td>Oct. 2015</td>
<td>Evaluation report completed and submitted</td>
</tr>
<tr>
<td>Nov. 2015</td>
<td>Focus group with evaluation team and project staff</td>
</tr>
</tbody>
</table>
than participants were available to enhance the range of possible interpretations, while paired and whole group conversations promoted inclusion.

The second activity used a placemat strategy to document perspectives. Groups of four participants (two members from the organization and two members from the evaluation team) completed a “placemat.” A placemat is a template designed to facilitate collaborative thinking and recording focused around a specific prompt. After discussing key ideas that emerged, the focus group continued with dialogue structured as questions and answers. At the close, all participants were invited to write confidential comments or questions for the focus group facilitators.

Data Analysis

All interviews and the focus group discussions were recorded and transcribed verbatim. After completing an initial read-through of the data, we created a draft analysis template using the eight principles as a heuristic. Space was also identified for emergent ideas. Initial coding for a single principle was subsequently undertaken by two of the authors who also compared codes to check for disagreements. While disagreements were few, the ones that did exist were resolved through discussion. At the root of these disagreements lay the acknowledgement that a particular portion of the data could be appropriately used for more than one code. Once this process had been completed for a single principle, it was repeated. Again, agreement was excellent, so the remaining data were coded independently, with any disagreements being resolved at the end. We see this overlap and dialogical process as a strength.

Once all data had been coded, we completed analytic memos for each of the principles (Center for Evaluation and Research, 2012). The memos were reviewed and edited by another member of the research team, until a final set of memos existed. These memos served as our final product of analysis and were reviewed with the original template to formulate the findings for this article. Although there were data and analytic ideas to support each of the principles, during our re-reading and discussing we decided that only a selection of the principles could be comprehensively covered in reporting this study. To this end, we selected the four principles that had independent and interrelated data.

FINDINGS AND DISCUSSION

Each of the CAE principles we studied in this evaluation context made a different contribution to our thinking. Our primary consideration in focusing on four principles was one of space and readability of this study. We wanted to provide an in-depth examination of the principles, rather than a cursory overview. Readers interested in our coding structure or data on the remaining principles should contact the first author. In selecting the four principles, we were driven by the focus of our study, the results from our analysis, as well as the distribution and interrelationship of the ideas brought forward from the data. The principles we examine are (a) fostering meaningful interprofessional relationships, (b) shared understandings, (c) evaluation progress and quality, and (d) evaluative thinking. We
offer a brief theoretical framing of each of these principles followed by examples illustrating how these principles played out within our DE context.

**Fostering Meaningful Interprofessional Relationships**

Shulha et al. (2015) identified successful collaborative approaches as dependent on the “quality of the inter-professional relationships that evaluators and clients/stakeholders are able to develop and sustain” (p. 17). Relationships have long been established as an important aspect of program evaluation (e.g., Abma, 2003, 2005, 2006; Abma & Widdershoven, 2008; MacDonald, 1977; Patton, 2000, 2002, 2007, 2008). At the core of this literature rests the idea that evaluators need contact with people to understand programs and their effects (Kushner, 2000). When using CAE, the importance of relationships is amplified. Three interrelated concepts emerged in studying this principle: (1) relationships evolve between people over time, (2) there is a layering of relationships, and (3) relationships emerge in response to people and the processes situated in particular contexts. Each of these concepts is examined in this section, and the overall value of this principle is discussed later.

*Relationships as evolving.* Relationships evolve over time and comprise many qualities. Relationship qualities that emerged as essential included clarity of roles, trust, and credibility. At a minimum, everyone participating must have a clear idea of what the program and the DE are designed to achieve. Yet we discovered that in this evaluation these essential qualities were initially defined by uncertainty. Furthermore, because of the CAE in this DE context, the uncertainty did not impair the overall interprofessional relationships.

From the outset, there was uncertainty about the goals, scope, and even the nature of the pilot project. For example, there was uncertainty about the roles and relationships within the organization as well as the role of the external evaluators. Although inevitable in DE, this uncertainty affected the relationships amongst group members. On one hand, it was positive for people to jointly enter into the work early in the process. On the other hand, the uncertainty of the project and relationships was compounded by the selected evaluation approaches. Many organizational participants commented on this instability during the midpoint interviews, reflecting on it again during the focus group session. As one participant noted, “we had some wrestling of what exactly is the role that everyone is playing here.” Another participant stated, “I think the challenge was really understanding what the role was within the larger scheme of what we were doing.” Participants commented on the overall need for deeper understanding of the people involved, the project, and the evaluation purposes.

In spite of an initial lack of clarity, the commitment, expertise, and openness of that group fostered strong interprofessional relationships. One participant commented on the importance of developing relationships early:

> I think one of the differences when you come in early is that relation piece. When you come in at the end, the program has already been developed. Those relationships are starting to change internally and it’s hard to come in externally and build that internal-external connection.
Many participants commented on the different lenses, the complementary expertise, and the usefulness of joint-critical thinking. As one participant said, “every step of the way there has been some input . . . it helped to spark our conversation.” While there was recognition of the value of this, participants at the final focus group also recognized that there may have been missed opportunities to utilize or maximize possibilities for learning. For example, one participant commented, “whether we’ve capitalized on that [expertise] the way we could I’m not sure.” These data underscore that developing interprofessional relationships in CAE is a process; learning to reap the rewards is also a process.

The ability to constructively use the evaluation information and thinking was enhanced as the relationships deepened. The initial trust provided a foundation of collaboration and credibility as the project proceeded. As one participant shared her experience:

I think the thing . . . is the fact there was a relationship that was built and that trust . . . there was this unknown, of who everybody was, what kind of experience did they bring to the table, strengths, weaknesses so on.

Participants commented that as interactions and their understandings evolved, they gained perspective about the project, the role of DE, and CAE.

We discovered that interprofessional relationships were enhanced by a conscious effort to attend to factors that allow the collaboration to grow in its effectiveness (e.g., trust, explicit valuing, and role clarification). The following image and narrative were selected by staff from the organization to reflect the quality of the relationships between the evaluation members and the organization.

The lead evaluator had worked with the organization previously and was known and respected for this past work as well as her work within the evaluation community. This prior relationship set the tone for a meaningful experience. Using the same image to reflect, one organizational participant described, “I think the fact that there was such a large team all coming with so many different interests, backgrounds, perspectives, perceptions . . . ultimately got us rowing in the right direction.”

Relationships as layered. Multiple relationships were at play within the DE context. For example, there were internal, full-time and contracted organization staff, external evaluators, and both English- and French-speaking participants from a range of school districts. Central to these different relationships was the interplay between our roles as a group engaged in a project. As one organizational participant explained, “what I ended up with. . . is the main realization that this is a collaboration on multiple levels.” The collective group was made up of a network of cooperative relationships and the web formed by these relationships lent strength to the collective.

Another organizational participant described how “people were talking, engaged, formal, peer-to-peer . . . they are sharing and learning through sharing.” The evaluation members and organizational participants enacted various roles: as coaches, supporters, facilitators, critical friends, helpers, and audience members who supported learning. This work was conducted in person and also used an array
of technology-mediated platforms to overcome distance barriers. A participant explained, “some of the greatest relationship development was through the one-on-one. Conversations with […] in the car ride with […] and […] where we really got to unpack the thinking and ideas and contributions going into the project.” Organizational participants valued both the formal (e.g., retreats, summaries of data, analytic memos) and informal (e.g., lunch chats, telephone calls, and e-mail) modes of communication, especially when the informal communication was face to face.

Relationships as emergent. Organizational participants also spoke of recognizing that there were iterative cycles to our work together. These cyclical processes provided opportunities for relationships to emerge and deepen. An organizational participant suggested, “there are stages, it is process, it is not just one thing. It is not an isolated event or activity.” Joint cycles of work relates to the concept of interactivity, which links the process of participation with enhanced potential for evaluation use. As Huberman (1999) suggested, through sustained interactivity we can refine our conceptual frameworks to enhance understanding. Although the project has ended, the interprofessional relationships can morph and continue to emerge as meaningful. In fact, several organizational participants identified friendship as an important outcome of the CAE project. Interprofessional

“We came in with a little bit of a history right? So we had a little bit of trust going there. And so over time what happened, what I thought happened, what was there was a trust amongst the whole group. That we began to trust each other and somewhat like what happens when you are both in a canoe or in a rowboat you learn when you have to pull a little bit stronger to keep the thing going straight.”

Figure 1. Rowboat image used as a prompt in the post-evaluation focus group.
relationships allowed differing viewpoints and areas of expertise to be valued. We
discovered that collaboration emerges within and across multiple roles/groups
and is not linear; one participant likened this experience to a journey.

In this principle people are engaged in talking, sharing, reviewing, question-
ing, and generating ideas—activities that emerge in different contexts in ways that

Figure 2. Path image used at focus group
are both intentional and unintentional. This principle demonstrates that valuing others, promoting respect, honesty, trustworthiness, and transparency is a process that unfolds as relationships deepen. Each of these qualities is significant, and can be developed over time with attention, skill, and effort.

**Developing a Shared Understanding of the Program**

Shulha et al. (2015) cite evaluation standard A4, “explicit program and context descriptions” (Yarbrough et al., 2010), as support for developing a shared understanding of the program. However, in a DE context, programs are in an emergent state, and concrete descriptions of programs often unavailable. Therefore, it is necessary to move “beyond the elements of a logic model” (Shulha et al., 2015, p. 11) and develop a shared understanding of not only the program logic/goals, but also the goals of the DE process. The process must include consideration for the roles that different groups of people could play in fulfilling those project/program goals.

While not unusual in dynamic evaluation contexts, our findings revealed that a shared understanding developed over time in a constructivist fashion (Lincoln, 2003). Arriving at the project model was neither straightforward nor obvious at the outset. Rather, it was the result of systematic and ongoing assessment of the areas where the project could have the most impact. For instance, at the beginning the organization was uncertain as to which means of delivery would be most effective for facilitating teacher professional learning, or what outcomes were reasonable for such learning given the context and time frame within which the project was likely to unfold. Evaluating the project using more conventional approaches would have required the imposition of a set of standards, expectations, and assumptions, constraining the evaluation as the program evolved and changed. The focus group data point to the impossibility of developing a shared understanding of the program until the goals are solidified. Even more interesting was the considerable evidence suggesting that members of the organization saw DE itself as an integral part of the project. A shared vision thus came to include both the project and the evaluation.

The evaluation members recognized that CAE provided opportunities to reframe conceptual, operational, and logistical challenges and “missteps” as positive learning. Adopting CAE within a DE allowed the group to take an evidence-informed approach to guide ongoing project development. For example, ideas and strategies were refined over time, and project activities became more targeted. At the closing day, rather than simply having teachers share their learning from the project, this sharing was constructed as a data-capturing activity. Seeing the data potential in the activity may not have happened without the CAE in this DE. Members of the organization were afforded multiple opportunities to “reality-test” their assumptions and generate rapid feedback to inform decision-making.

A key element proved to be defining the goals and structures while also sorting through the role that evaluation would play. There were two main reasons for this. First, the project was multidimensional, so that in addition to developing a project, the organization also wanted to build evaluation capacity. Second, the funder’s
expectations and requirements changed throughout the project. These two factors likely explain why uncertainty permeated organizational members’ comments about DE at the beginning of the project. When the focus group facilitators summarized the discussion of DE at the beginning of the project, there was much laughter as everyone realized their level of anxiety and uncertainty was shared.

The evaluation members made an effort to explain DE from the outset, but it appears that didactic approaches were not effective and that participants needed to go through the process to fully understand it. As one organization participant said, “in spite of the fact that I know we had a session in which DE was presented to us, I think the . . .experiential, hands-on minds-on, of actually being immersed in it helped to shape my understanding.” Even though participants from the organization weren’t always clear about project goals or how to achieve them, they were willing to keep moving forward. This ambiguity seems to be a hallmark of the types of projects for which DE is designed and implemented (Patton, 2015).

There were ample data about developing a shared understanding of the project, and this suggests to us that, next to fostering relationships, the program itself is central. In this case, participants from the organization recognized that the beginning of the project lacked direction, and looked to the evaluation members for assistance. Yet, when guiding questions were provided instead of concrete advice, participants’ frustration was enhanced. One participant illustrated this at the mid-interviews by saying they wanted “more leadership as opposed to here's things to think about based on what we heard you say.” We learned that the frustration reached a high level when a lead from the organization revisited the evaluation agreement to ensure the evaluation members were fulfilling their obligations.

It was only in retrospect that the client organization realized “you couldn't focus on exactly . . . what we needed because we weren’t clear on what we needed.” It is clear that in a DE context, program models are going to be emergent, dynamic, and unstable. As a result, evaluators can use CAE to work through ambiguous and evolving program models. Since both project development and evaluation can induce or enhance frustrations and anxieties, evaluators can rely on strong interprofessional relationships to maintain positivity and momentum.

**Monitoring Evaluation Progress and Quality**

One of the important themes of this principle is that of flexibility and adaptation to change. Shulha et al. (2015) wrote, “it is not unusual for the evaluation context (i.e., people, programs, organizational context, information needs, and process needs) to change” (p. 15). The literature about DE has established that change is expected, and so the evaluator must manage change skillfully (Patton, 2011b). Monitoring evaluation progress and quality during change requires a complex set of professional attributes and vigilance.

Evaluators are responsible for ensuring that processes and products are of high quality. Guiding our notion of quality are standards of evaluator professional practice put forward in the *Program Evaluation Standards* (Yarbrough et al., 2010). Evaluation quality can also be monitored based on the professional domains, which
highlight the competencies established by the Canadian Evaluation Society (http://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf). Both of these professional standards and competency domains were embedded in the work of the present evaluation. Lyn Shulha is both an author of the Program Evaluation Standards and the lead of this evaluation team. Moreover, she and the other members are active in professional associations, which have shaped their understandings of evaluation quality. In this case, an understanding of monitoring evaluation progress and quality was established right from the start when the evaluation team introduced the idea of monitoring our evaluation practice as research.

A second theme that Shulha et al. (2015) highlight for the principle of evaluation progress and quality is that of data integrity. Data integrity refers to assuring and maintaining the completeness, accuracy, and consistency of the data. In this study we found that data integrity was inextricably linked to building evaluation capacity and fostering strong interprofessional relationships. Having participants from within the organization collaborate with the university-based evaluation team led to an array of perspectives about what counts as data and how data can produce information that contributes to different types of use. It is clear from the focus group data that participants saw this group diversity and collaboration as critical components of a quality evaluation. In response to the image of a rowboat presented earlier in this article, one organizational participant stated:

I think the fact that there was such a large team all coming with so many different interests, backgrounds, perspectives, perceptions . . . ultimately got us rowing in the right direction and bringing our perceptions to the table to drive the project, [but the large team] also [made it] hard to get started.

In other words, the same diversity that contributes to quality also poses challenges to the achievement of a shared vision. It is interesting to note that having diversity among the members of the evaluation or organizational group is not a theme specifically addressed by Shulha et al. (2015). Diversity emerged as a prominent theme in our findings across multiple principles, including the principle of evaluation progress and quality. The program evaluation standards require evaluators to “devote attention to the full range of individuals and groups invested in the program” (Yarbrough et al., 2010 p. 3). This seems important in CAE, especially when nested in a DE. In this case, CAE fostered a space for the interrogation of ideas and choices that enhanced the quality of the work.

Our data further revealed that formal checkpoints for monitoring evaluation quality served to reflect on evaluation progress and promote the development of the organization's capacity for evaluation. One of the facilitators of the focus group session summarized the discussion surrounding evaluation progress by saying:

What I am hearing from everyone is that it is one of those concrete moments, when we got to the planning days, or had the interviews, or when we did the report writing or joint analysis, some of those concrete tasks that we engage in as part of our process are emerging as critical.
The placement activity provided a structure for participants to reflect on their experience and progress in the project and the evaluation. While the focus group activities led to good-natured groaning, they also helped to target participants’ thinking. During the project, quick “check-ins” from the evaluation team were critical nudges that helped move the project and evaluation along.

Formal checkpoints also served as opportunities for the organization to ask questions of the evaluation members. As project developers prepared for the checkpoints, they raised questions about the evaluation's progress, what data needed to be collected, and what types of analyses would yield the most useful information. During the midpoint interviews one participant asked for help with designing a data collection instrument, explaining:

I would like some support and help [with the] final survey . . . I know the kinds of information that I would like to gather but in terms of structuring the questions around it . . . I would certainly appreciate your input into that.

This is evidence of evaluation progress and quality, leading to enhancement in evaluative thinking. This participant reflected on what he or she wanted to know and the skill set he or she wanted to develop while looking to the evaluation members as partners in evaluative knowledge building. The increased evaluation capacity that is articulated here came about, in part, because the evaluation members deliberately designed and selected checkpoint activities to enhance process use and evaluation thinking. As one would expect in CAE, the evaluation members were continuously attending to stakeholder needs, monitoring to make sure these efforts were working appropriately. These reflections ultimately led to processes such as the use of image-elicitation to augment the focus group. By using these activities, rich data were collected that promoted programmatic and evaluative understanding. King et al. (2007) describes this as “planning intentional process use” (p. 47). Monitoring evaluation process and quality promotes reciprocal learning and enhances evaluation practice. In this CAE the principle is a reminder of the importance of being responsive to change, creating opportunities for reflective learning, and meta-evaluation.

**Promoting Evaluative Thinking**

An attitude and belief in the “value of evidence,” driven by evaluators and participants who foster skills such as critical thinking, creative problem solving, and questioning in collaborative contexts is frequently referred to as evaluative thinking (ET). Shulha et al. (2015) recognize the need to cultivate ET within CAE. Buckley, Archibald, Hargraves, and Trochim (2015) offer the following definition:

Evaluative thinking is critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action. (p. 378).
Promoting ET embraces openness to the unexpected so that programmatic and evaluative discoveries can lead to an ongoing investment in learning. The goal of this orientation is to encourage informed decision-making and propel action (Archibald, 2013).

A review of the related literature revealed the ensuing four key dimensions of ET: (a) intentionality, (b) practice based, (c) challenging assumptions, and (d) intrinsic and incremental (e.g., Amo & Cousins, 2007; Baker, 2011; Baker & Bruner, 2012; Carden & Earl, 2007; Davidson, 2005; King, 2008; Preskill, 2008; Preskill & Torres, 1999). These dimensions are conceived in complex and dynamic ways depending on the context and the readiness of the team to engage in evaluative thinking. There is overlap with the principle put forward by Shulha et al. (2015); we examine how our data support these dimensions.

**Intentionality.** Promoting ET begins with the explicit value for evaluation that is established, sustained, and/or cultivated. One participant noted that evaluation processes are “generally not valued . . . there are a few of us that really think it’s important but probably not enough.” In response to this belief, participants talked about the capacities they developed as individuals or as a team through this project. In this case, promoting ET was about embarking on a process of concurrently learning about one another, funder requirements, school district requirements, organizational requirements, the perspectives and skills of those involved, the pilot project itself, and DE.

By intentionally positioning ET as part of CAE, evaluators prioritized learning throughout the evaluation. One participant described evaluation as “unsettling, and not everybody likes to be unsettled.” Yet, participants identified value in intentionally disseminating the knowledge and skills gleaned from this project for the betterment of the organization as a whole. As one participant said, “we need to be able to impart this knowledge [to] others!” ET supports capacity building at individual levels; it can be infused into programs and organizational thinking. As participants talked about how they would transfer knowledge, they also made some sophisticated and pithy comments about conducting quality evaluations. One participant recognized an important limitation of data collected in a dynamic context:

You capture data in a moment in time. Data is what keeps the project going, but data is still. You gather data from a survey on a particular day. You know you capture people’s perspectives to date on that particular thing. Then you do it again. So I feel like data it is a snapshot of a program, it is not the program.

Another participant revealed her increasingly complex views on data collection by asking the simple question, “What are the types of questions that serve us to elicit the feedback that we need to know?” These are examples of learning that can occur when ET is intentionally cultivated in organizations through CAE.

**Practice based.** ET is stimulated in the contexts where the inquiry takes place. Developing an evaluative mindset requires individuals to ask questions without perceiving questioning as an attack on personal performance or program development. A participant used the metaphor of developing a thick skin to illustrate
the change in attitude required. The reference denotes an understanding that growth or improvement is often difficult and requires focusing on programs and professional practice. In this case, one participant described how the evaluators offered “information to digest” and “clarity within our processes.” A focus on ET in practice-based contexts stimulates deeper learning by enabling teams to move between program and organizational learning.

**Challenging assumptions.** Most participants characterized their initial thinking using words about assumptions, muddied understanding, misconceptions, and even erroneous ideas about evaluation’s potential contributions. In our midpoint data, significant conceptual confusion surrounding the difference between evaluation and other forms of inquiry was revealed. Distinguishing evaluation from research has been a focus of theorizing in the field of evaluation for some time (e.g., Fain, 2005; Fitzpatrick, Sanders, & Worthen, 2004; Levin-Rozalis, 2003; Patton, 2014). Some participants attributed their initial thinking to past experiences where research was viewed as more esoteric and evaluation as a top-down activity that did not easily generate useful findings. Yet, as the collaboration progressed, participants reported that they were able to see how DE promoted thinking and was directly applicable to practice. This finding is typified in one participant’s comment: “you cannot develop your program activities independent of what you want to learn.” When you challenge assumptions and overcome pre-existing thinking, you provide a pathway for deeper learning.

**Intrinsic and incremental.** Our data also revealed how ET develops over time. The value of CAE was described as a success by a participant: “If you are embarking on a project . . . or have an outside perspective . . . listening to the discussions and providing suggestions... [They] challenge you to make sure that the program developers are thinking of this really in a fulsome perspective.” As the project progressed, the evaluation members used findings to promote thinking and programmatic decision-making. As one participant observed, the evaluation group was clear about providing information as opposed to telling us what had to be done. Another participant noted, “your feedback is often ‘things we might want to consider based on what you’re reading’ that is different than ‘this is what the information is’.” One of the goals of CAE in a DE is to stimulate thinking in incremental ways. This principle underscores the importance of staying focused on the big picture even while attending to the inevitable details that arise in program evaluation. Overall, the principle of *evaluative thinking* speaks to an orientation that values learning as a process (Shulha et al., 2015).

**CONCLUSION**

Evaluators use a range of methods to promote continuous learning. Reflective practice is one such technique, enabling evaluators and evaluation scholars to consider the thinking, assumptions, frameworks, and patterns that shape their actions. Reflexivity is a critical element when using participatory approaches (Heron & Reason, 1997). This requires purposefully locating oneself within the
research process to minimize the unintentional effects of power (Hesse-Biber & Leavy, 2010). The value of reflective thinking is highlighted in the Program Evaluation Standards (Yarbrough, et al., 2010) and is considered one of the six essential competencies as identified by Stevahn, King, Ghere, and Minnema (2005). Given the dynamics of politics and power relations in evaluation, self-reflexivity is critical to creating transparent processes and learning from these processes. Reflection is at the core of Lyn Shulha’s practice, and it is the heart of this work.

In this study, we engaged in reflective practice at two points for the purpose of learning about the particularities of the complex, nuanced principles of CAE in DE. Active, ongoing, and critical reflection of evaluation processes illustrates the importance of inquiry activities for stakeholders throughout a project, not just at the end. We have examined four CAE principles as they played out across our midpoint and endpoint reflections. Intentionally structuring opportunities for and using explicit reflective processes during CAE in DE can deepen understanding about programs and evaluation, build professional relationships, and enhance group communication. We recognize that we have more to learn. Next, we will write up our findings of the remaining principles, while also looking for future sites for empirical study of the CAE principles. In addition, a future study could examine which principles lend themselves better for the beginning or midpoints of the collaboration or later stages of the program development or implementation. As Shulha et al. (2015) indicate, the decision about which principle to emphasize and when, is likely 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 the complexities as the evaluation unfolds. (p. 12)

We chose to emphasize the four principles where we had the greatest volume of independent and interconnected data so that our unpacking of these principles could contribute to deeper understanding of their application in practice and to the field more broadly.

Reflective thinking about the CAE principles in this DE provides the following lessons:

1. Fostering meaningful interprofessional relationships sounds like a straightforward idea, but relationships are inherently complex. Collaborative approaches are about relationships. Within this principle is embedded the need to provide a service to users while also maintaining professional integrity. Relationships among stakeholders and evaluators do not exist in isolation, as they are interrelated and influence one another.

2. Developing a shared understanding of the program within a collaborative approach requires attentiveness and responsiveness to the unique, complex, and challenging situations that exist and arise in all of our professional work.

3. Monitoring evaluation process and quality when undertaking collaborative work in DE contexts requires a shared vision and commitment that
is sustained through formal and informal processes, which are enacted in explicit and implicit ways through sustained interactivity.

4. Promoting evaluative thinking relates to the need for evaluators to be responsive to change, creating opportunities for reflective learning and meta-evaluation.

Using the principles to reflect on lessons learned from this evaluation context demonstrates the complexity of enacting collaborative work in DE. These principles may be of particular importance in DE where the evolution of a program can prompt reflection and joint opportunities for learning. The CAE principles provide a set of lenses that allow the reflective evaluator to focus on the elements of collaboration, allowing her to better understand the challenges and success of the evaluation.

In a letter to the editor, Fetterman, Rodríguez-Campos, Wandersman and O’Sullivan (2014) disagreed with the ideas initially put forward about the creation of the CAE principles. Yet, Cousins, Whitmore and Shulha’s (2014) response refers to a subsequent plan involving those same authors to develop a set of unifying principles for “stakeholder-involved” approaches. The issue of terminology, similarities/difference, and conceptual clarity is an ongoing dialogue and debate in which we are interested to take part. A key distinction is that the CAE principles were developed through empirical inquiry and are used in this study in an applied setting as a way to understand evaluation practice from the perspectives of the evaluators and participants from the organization. Lyn Shulha is to be credited with many of the foundational ideas in this article and the reflective nature we enact in our evaluation work. She has instilled in us the value of enhancing our scholarship in the area of program evaluation through application of evaluation theories, and ongoing development of the evaluation standards and competencies of the field. Our collaborations have taught us that there is great value in having mentors who are both scholars and practitioners as well as working with practitioners who are also versed in scholarship. The principles presented in this article remind us that we need to be continually reflecting and refining our thinking about CAE in different evaluation contexts and always be attentive to what we can learn from the intersections of evaluation theory and practice.

REFERENCES


Guiding Principles for Collaborative Approaches to Evaluation


AUTHOR INFORMATION

Michelle Searle holds a PhD from Queen’s University and works as an independent educational researcher and evaluator as well as an instructor at both Western University and Queen’s University. Her work is focused on educational evaluations that promote learning with multiple stakeholders. She draws from her background as an arts educator to interweave creative practices into program evaluation contexts.

Stefan Merchant, PhD(c), is currently completing his dissertation in educational assessment at Queen’s University in Ontario, Canada. The focus of his research is the assessment of complex constructs such as learning skills and work habits. His interest in program evaluation stems from his experiences as an educational administrator.

Agnieszka Chalas, PhD(c), is a doctoral student in the Faculty of Education at Queen’s University whose research focuses on examining program evaluation practice and capacity in the art museum sector. Her research interests are informed by over ten years of senior management experience in the arts and culture sector where she assumed broad responsibilities including internal program evaluation.

Chi Yan Lam, PhD, CE, is a credentialed evaluator and educational researcher. His current program of research examines the use of developmental evaluation and the role of design theory in promoting innovative development of programs.