Reflections on the Meaning of Success in Collaborative Approaches to Evaluation: Results of an Empirical Study

Elizabeth Whitmore
Carleton University

Hind Al Hudib, J. Bradley Cousins, and Nathalie Gilbert
University of Ottawa

Lyn M. Shulha
Queen's University

Abstract: *How do evaluators using collaborative approaches to evaluation (CAE) define success? This is the core question being asked in a further analysis of data from our previous work (Cousins, Whitmore, & Shulha, 2013; Shulha et al., 2016) that developed a set of evidence-based principles to guide collaborative evaluation practice. Probing data from 320 responses to our (2012) survey, we examined what respondents considered “highly successful” and “less successful than hoped” in their collaborative evaluation projects. The results revealed that evaluation use, relationships, and information needs are key factors. We propose a conceptual framework as an aid to thinking about success in CAE.*

Keywords: *collaborative approaches to evaluation, success*

Résumé : *Comment les évaluateurs utilisant des approches collaboratives à l'évaluation définissent-ils le succès? Voici la question de base posée dans une analyse plus poussée de données tirées de travaux précédents (Cousins, Whitmore et Shulha, 2013; Shulha et al., 2016) qui ont permis d'élaborer un ensemble de principes scientifiquement fondés visant à orienter la pratique de l'évaluation collaborative. En examinant les données de 320 réponses à notre sondage (de 2012), nous nous sommes penchés sur ce que nos répondants ont jugé être des projets d'évaluation collaborative « très réussis » et « moins réussis qu'espéré ». Les résultats ont révélé que l'utilisation de l'évaluation, les relations et les besoins en information sont des facteurs clés. Nous proposons un cadre conceptuel pour penser le succès en matière d'approche collaborative en évaluation.*

Mots clés : *approches collaboratives en évaluation, réussite*

Corresponding Author: Elizabeth Whitmore, 55 Second Ave., Ottawa, ON K1S 2H4;
Elizabeth.whitmore@carleton.ca

INTRODUCTION

In this article we are interested in furthering our understanding of what it means to be successful in collaborative approaches to evaluation (CAE), evaluations where trained evaluators work in partnership with members of the program community to produce evaluative knowledge. CAE is a class of approaches to evaluation that are alternative to mainstream or conventional options. As is the case with much of evaluation practice, we need to further our practical and theoretical understanding of CAE through empirical inquiry (Cousins, Whitmore, & Shulha, 2013). In this study, through secondary analysis of data, we drill down into the concept of “success” in CAE. First, however, we review what we know in more general terms about the concept.

Initially in evaluation discourse, program success was generally equated with having achieved the program goals. However, the literature reveals that this is a misconception about evaluation. Considering this, Scriven (2016) in his recent work begins a commentary on some key misconceptions about evaluation that act as roadblocks to the future directions that evaluation should take. One of the relevant examples that he identifies is the classic dictionary definition of summative evaluation: “an attempt to assess the overall effectiveness of a program in meeting its objectives and goals after it is in operation (p. 28).” He describes this as “completely wrong,” noting that, by this definition, Nazi prison camps would have scored well—“at least for several years” (p. 28). Scriven goes on to argue that focusing evaluation only on effectiveness (i.e., goal achievement) ignores key social obligations of a social science professional (such as goal critique and resource conservation) as well as side effects. In articulating this perspective, Scriven provides a welcome caution to evaluators who tend to focus uncritically on program goals as the only or primary indicator of success. At the same time, Poulin, Harris, and Jones (2000) argue that, although “an understanding of goals or definitions of success alone does not provide the entire picture of what a program tries to do or how a program functions, it does yield some important context for program evaluation” (p. 531). We can see that the concept of program success seems to be inextricably linked with notions of evaluation success.

So, what does program success mean and how is it reflected in the literature? Poulin et al. (2000) demonstrate that the definitions of success in a program can change over time and that documenting such changes can facilitate program development and policy making. Birckmayer and Weiss (2000) explore the role of theory-based evaluation in assessing a program’s success or failure.

These perspectives touch on a prominent theme, spanning many years of evaluation, which is the question of use—most often of evaluation findings. Though occasionally discussed in terms of “success” (Cousins, 1995), the assumption has been that if evaluation findings have been used or determined to be useful by the evaluand—in terms of intended use by intended users (Patton, 1978, 1997)—it follows that evaluation success has been achieved. Others, using words like “effectiveness” (Elbaz-Haddad & Savaya, 2011; Liket, Rey-Garcia, & Maas, 2014; Raphael & Stoll, 2006), or “what works” (Davies, Nutley, & Smith, 2000), imply that the ability

of evaluation to demonstrate such use reflects evaluation success. Some scholars have more directly referred to the usefulness of evaluation findings to policy makers and practitioners (Granger & Maynard, 2015; Liket et al., 2014; Wimbush, 2014) or for organizational learning (Cousins, 1995; Fetterman, Kaftarian, & Wandersman, 2015; Preskill, 2014; Torres & Preskill, 2001) as indicators of evaluation success. Still others focus on “stakeholder engagement” as the indicator of success (Adams et al., 2015; Brandon & Fukunaga, 2014; Liket et al., 2014; Sturges, 2015). Fetterman et al. (2015), in their work on empowerment evaluation, emphasize the capacity building of stakeholders so that they have the logic and tools to “plan, implement and evaluate their own programs” (p. 2). The most popular empowerment evaluation models that have emerged over the many years of dialogue and development of the approach—notably the 3- and 10-step models—“enhance the probability of program success” (p. 9). Of interest is that virtually all of these contributors embrace a positive psychology perspective in thinking about success. What does it mean to be successful? What does success look like?

Yet, another way of learning about what evaluation success means is examining less-than-successful examples. This was the topic of a 2010 special issue of the *Canadian Journal of Program Evaluation* titled “As I Recall—Or How to Take Advantage of Less-Than-Successful Evaluation Experiences” (Gervais & Joubert, 2010). While most writing highlights best practices, this collection of case studies examines “worst practices” so that we can learn from our mistakes and use this information to improve practice. Common themes in this issue include communication, the importance of standards and evaluator competencies, and engagement with key stakeholders.

Whether looking at the glass as half full (successful) or half empty (less-than-successful), the importance of good and ongoing communication is seen as essential: “When communication is not good, it presents serious threats to the success of an evaluation” (Connor, 2010, p. 128), yet “good communication facilitates but does not guarantee success” (p. 133). The concept of “relationship” is another vital factor. “The key to a successful or unsuccessful evaluation is often the quality of the relationship between evaluators and their clients” (Hawkins, 2010, p. 27). This implicates the interpersonal skills of the evaluator, called “soft skills” by Perrin (2010) and “people skills” by Patton (2010), such as “negotiation, conflict resolution, collaboration and diversity” (Patton, 2010, p. 156). Patton emphasizes the issue of complexity, noting the interplay of stakeholder issues, contextual factors, and evaluation management practicalities. He argues that this entails far more than simply constructing a list of success factors; rather it “involves understanding the complex dynamic interactions among those factors, dimensions, and competencies” (Patton, 2010, p. 158). Owen (2010) concludes that success is a relative concept: “what might be regarded as success for one involved party might not be regarded as such by another” (p. 86). Program advocates, as an example, may consider an evaluation to be “unsuccessful” because it did not support their views, while an evaluator, attending to the integrity of the process, felt that the same evaluation was successful in that it adhered to professional

standards of practice. Rather than thinking only in binary terms (i.e., successful/not successful), as [Patton \(2010\)](#) suggests, it may be more constructive for us to think in terms of “incomplete successes” and to acknowledge that learning from these can “teach us how to evolve and adapt” ([Gervais & Joubert, 2010](#), p. xvii).

In a separate publication, [Wandersman \(2009\)](#) discusses four keys to success in participation, with one essential element being the disposition of the evaluation itself. He suggests that “evaluation failure” can occur through poor design, inappropriate measures, or negative stakeholder experiences and therefore attitudes toward evaluation.

Some evaluators propose tools or methods as being helpful in determining success in programs and other interventions. These contributions are of interest to the present discussion about evaluation success because they provide insights into the success construct. For example, [Marek, Brock, and Savla \(2015\)](#) outline a collaborative assessment tool (CAT) used in evaluating the success of collaborative program efforts:

As evaluators are increasingly asked to evaluate collaborations and coalitions, this conceptual model and tool can provide evaluators with a grounded, reliable, and valid assessment instrument to work with clients to build collaborative efforts in an intentional, comprehensive, and effective manner. (p. 67)

In a similar vein, [Mills, Crone, James, and Johnston \(2012\)](#) used a mixed-method design to highlight the multidimensional nature of success in exercise referral schemes. Their design “broadened the focus beyond physical outcomes, to include psychosocial factors associated with behavior change[, providing] a better understanding of success” (p. 421). What is important for the present discussion is that they found that success is not a static concept; rather, the perceptions of success have the ability to adapt over time, as stakeholders experiences change. [Brinkerhoff \(2003\)](#) describes a method—the success case method—that looks for successes in an initiative, even if some or all of these successes are partial. This method uses “persuasive or compelling stories” that lead to “a better understanding of why things worked, and why they did not. With this knowledge, success can be built on and extended; faltering efforts can be changed or abandoned, and promising efforts can be noticed and nurtured” (p. 1).

Success in particular domains is explored in other related literature. For example, [Walter and Scholz \(2007\)](#) highlight critical conditions or factors that are necessary for success in collaborative planning projects (in this case, in urban transportation). [Sawhill and Williamson \(2001\)](#) present a model for measuring success in not-for-profit organizations (using an example of Nature Conservancy). Finally, [Moehr et al. \(2006\)](#) look at success factors in telehealth.

The foregoing scholarship helps us to understand the construct of success not only in terms of evaluation but also from the perspective of social and other interventions, thereby providing valuable insights into its nature and essence. One conclusion that is clear from our review is that empirical inquiry regarding evaluation success is really quite sparse. We observe, however, that successful evaluations

are identified not only in terms of their consequences (i.e., evaluation use) but also that process elements such as communication and relationship development are important. These elements seem quite relevant to CAE since, by definition, evaluators work together in partnership with program community members or stakeholders to co-construct evaluative knowledge.

In the present study, we were interested in how evaluators frame success in CAE. Specifically, we explored questions about their practice experience with CAE and we indirectly, somewhat inferentially, investigated their definitions of success. Based on the findings, we propose a conceptual framework as an aid to thinking about success in CAE.

METHODS

This exploratory study is based on the secondary analysis of data collected for another purpose, to develop principles to guide collaborative approaches to evaluation. We have now produced and introduced that set of principles (Shulha, Whitmore, Cousins, Gilbert, & Al Hudib, 2015, 2016).¹ In that report, extensive details are provided about the methods used to gather the data for the study. In this section, we describe how we went about conducting the secondary analysis of the data.²

Data Structure

In the main phase of our data collection, 320 practicing evaluators completed our online instrument that included quantitative and qualitative items. Participants were asked to identify a CAE project from their own experience that they considered to be highly successful and provide a set of responses about that project including two open-ended questions: (a) “What were the top 3 reasons why this collaborative approach to evaluation was highly successful?” and (b) “Provide more details about the project (e.g., purpose, context, other reasons).” Data associated with this supplementary open-ended item provide the principal focus for our secondary data analysis in this article.

It should be noted that the participants actually identified two projects about which to describe and share their views. The second project was one they considered to be “far less successful than [they] had hoped.” In the larger sample the order of successful and less-than-successful projects was counterbalanced to control for response bias (Shulha et al., 2016). In the current study, our main interest is in projects that were explicitly identified as being highly successful. However, as discussed in the literature review, projects that were far less successful than hoped might also provide interesting clues about how evaluators define success, and so we analyzed those responses as well.

Analytic Strategy

Our general analytic strategy was to code the data from the supplementary open-ended question using an emergent set of codes. This strategy allowed us to identify themes associated with how evaluators define CAE success. It is important to note

that, by design, participants were not explicitly asked how they defined the success of CAE projects. Therefore, our secondary analysis of the data is necessarily limited by the extent to which participants decided to provide information relevant to the success of the CAE project. In some cases, they made explicit reference to project success, whereas in other cases we identified clues, hunches, and inferences worth considering and exploring. For this reason, it was important for us to also look at the reasons they gave for success, (i.e., data reported in [Shulha et al., 2016](#)).

Coding Structure and Data Quality Assurance

Two analysts (Al Hudib and Cousins) assumed principal responsibility for coding and analyzing the data using the qualitative data analysis software NVivo 10. After reviewing and discussing several responses and ideas, we generated a preliminary list of emergent codes and began to apply them to independent samples of the data. In doing so, we remained open to the possibility of identifying additional emergent codes. After having independently coded substantial portions of the data, we identified random segments of responses and coded them independently. We resolved identified discrepancies and then reviewed previously coded data and reapplied the codes accordingly.

[Table 1](#) displays the emergent coding structure and the frequency of application. It should be noted that any given response could be disaggregated into multiple ideas, and therefore multiple codes may have been applied. In fact, this was the case more often than not.

Three things are noteworthy about the contents of [Table 1](#). First, the emergent codes are broken into two categories, one corresponding to ideas about CAE project success (9 codes), and the other relating to extraneous details given by participants (4 codes). It can be observed in the table that the majority of codes actually apply to details about projects that are outside of our interest in success (e.g., background details about context, descriptions of evaluation purposes and processes). However, there are many responses that include rich information about dimensions of CAE project success, or at least clues and hunches about such phenomena. These latter responses are of high interest to our current analyses.

Second, we have sorted the nine emergent codes for success into a loose ordering corresponding to consequences or effects of the evaluation (i.e., use of findings, process use, benefits to the evaluator), stakeholder relationships and activities (i.e., engagement, intra-stakeholder relations, evaluator-stakeholder relations), and evaluation characteristics (i.e., purposes, resources, timing). Finally, we can see that the frequency with which codes are applied varied considerably over the nine codes. We will present the results associated with each code in the order in which they appear in [Table 1](#).

Data Reduction Strategy

After coding all of the data, we used NVivo to sort the responses into categories associated with the nine emergent codes or themes. In doing so, we created what we termed a “table of data elaboration” for each theme. These tables were

Table 1. Emergent Codes by Frequency of Application

Code	Description	# Code Applications
Defining Success/Nonsuccess		
Use-F	Use of findings or instrumental (program change), conceptual (learning about program), and symbolic benefits of the evaluation	25 moderate
Use-P	Process use or benefits to stakeholders (individual, team, organization) of the evaluation process independent of findings. Includes capacity building and transformational consequences of the evaluation	29 moderate
Relations-ES	Working relations between evaluators and stakeholders	63 high
Relations-St	Working relations among stakeholders	47 high
St- Engagement	Stakeholder engagement and participation in evaluation activities	32 moderate
Purpose	The evaluation aligned with intended purposes; stakeholder information needs met	40 high
Resources	Financial and human resources supporting the evaluation project	28 moderate
Evaluation Timing	Evaluation was timely in meeting stakeholder information needs	8 low
Evaluator-Ben	Benefits to the evaluator in terms of professional development, reputation, etc.	6 low
Other Details		
E-Purpose	Details about the purposes of the evaluation (not the program); no obvious connection to success (or lack of it) of evaluation	102 very high
E-Process	Details about evaluation process characteristics	32 moderate
Context	Details about context within which program (project, strategy, evaluand, etc.) resides	88 very high
Program	Details about the focal program and its effects. This would include unintended program consequences	119 very high

ordered by evaluator participant (rows), and included all data (columns) provided in response to the open-ended questions about (a) “other details” highlighting specific segments of text associated with the respective code and (b) “reasons for success” along with the associated codes that had been applied in the previous project (Shulha et al., 2015). These tables enabled us to develop a rich sense of

how evaluators were defining success in their projects, according to the respective dimension. We then summarized the results in the following section, theme by theme, including verbatim responses for illustrative purposes.

RESULTS

As mentioned, the nine emergent themes or dimensions defining successful CAE appear in [Table 1](#) along with frequency of occurrence (i.e., analysts' application or use of codes). We now turn to an elaboration of each theme in the order in which they appear in [Table 1](#).

Use of Evaluation Findings

To a moderate degree, our findings reveal that evaluators define the success of CAE in terms of specific evaluation consequences, that is, direct use of evaluation findings. We found many specific examples of *program changes* that were made on the basis of results, and/or the use of evaluation to influence program and policy decision making. Here is an example.

A follow up I did with the administrator the next year indicated that they had embraced the outcomes and were implementing suggestions that came from the findings. I believe the collaborative approach set the stage for the success of this project.

It is clear in our data that CAE feeds into an improvement process more so than accountability interests. Part and parcel of program improvement is “learning.” For example,

Participatory data interpretation process led to insights about program improvement that were immediately adopted and implemented, as well as deeper reflection about core issues in the philosophy of intervention.

Such deeper reflection is very much aligned with principles of organizational learning, with CAE serving here as a triggering event. But when thinking about the use of findings in terms of dimensions of CAE success, it is also important to consider the *accountability function*. Our results showed that adherence to accountability directives and compliance demands factored negatively into the conception of success. The evaluators spoke in terms of symbolic use of findings and its disingenuous qualities.

We also learned from projects described as less than successful that the *non-use* of evaluation data could be taken to define observed lack of success. We observed a political and/or nonrational element to this discourse.

Finally, we looked at what evaluators considered to be explicit *reasons for CAE project success* or lack of it in association with the “use of findings” success dimension. Of relatively high occurrence were considerations about the *relevance* of the evaluation to the information needs of the stakeholders. Other evaluators identified an organizational *culture* that is evaluation-friendly as being an

important reason for success. Another reason often related to lack of success in CAE had to do with the explicit *purposes* of the evaluation. *Interpersonal relations* were also important. One evaluator commented on the space that the CAE project provided for such relationships to develop: “The evaluation created opportunities to build trust between the relatively new administrator and other stakeholders.”

Process Use

Process use is a complex phenomenon associated with benefits arising from stakeholder proximity to the evaluation. This phenomenon was found to factor into evaluators’ conceptions of success in CAE to a significant degree. There were three specific aspects that emerged: direct or intentional evaluation capacity building (ECB), linkages with use of findings, and transformative effects.

In many ways CAE leads to the indirect development of evaluation capacity. Yet our data show that evaluators invested significantly in *intentional ECB*. Often these efforts led to projects that were highly productive and successful. The following is an illustration:

[The] focus was on providing evaluation capacity building and technical assistance to districts as they worked with their own program data. A great deal of time was spent early in the project to articulate program purpose and define roles. This was one of the most valuable steps to supporting the collaborative aspect of the project.

We observed that the learnings and skill development arising from participation in the evaluation are likely to be enduring and to transfer to future inquiry and other organization- or program-specific tasks. While process evaluation can and does arise naturally in CAE contexts, this may evidently be augmented considerably with direct and intentional ECB.

Process use is often framed as being independent of the *use of findings* but our data reveal a *linkage* between the two. Specifically, process use is mostly about learning, but the following examples illustrate that such learning may be interconnected with and even augment conceptual use or learning from the findings of the evaluation.

Discussions frequently went beyond the specific immediate and long-term outcomes to why the results occurred.

Examining the results together and thinking about what they meant helped them to recognize why the program did not work so well in some areas and how they might improve it.

Some of these responses reveal that CAE projects provide the space for deeper, more penetrating discussions about evaluation findings and their meaning.

There is strong evidence supporting a *transformative* aspect of the process use success dimension. Transformation takes the form of the development of relationships, organizational and program structures, and understandings about program and organizational capacity.

Both efforts (the process of developing the monitoring system and the program-specific evaluation TA activities) led to highly engaged stakeholders and exponentially improved our program's understanding and valuing of evaluation.

[The evaluation] allowed us to build a much more complex and horizontal web of relationships which promoted buy-in and feelings of accountability and responsibility in both directions.

Finally, we looked at the main *factors* influencing the process use success dimension and observed a wide array of reasons or factors. Among the more prevalent was *relevance*, defined mostly in terms of shaping evaluation objectives and enhancing receptivity to findings: "Made [the stakeholders] more receptive to the results" or "stakeholders determined the evaluation questions." But *depth of participation* in the evaluation by stakeholders was also highly influential as suggested by the following quotation: "all parties involved were highly engaged in this evaluation work and contributed extensively to the content of instruments."

Two additional factors or reasons interrelate with the foregoing. First the specific nature of the *purposes of the evaluation*, specifically with a focus on learning, was found to be important. The second had to do with stakeholder *information needs* and their specific interests in the evaluation and the findings it was likely to generate.

Evaluator-Stakeholder Relations

To a considerable extent, our findings reveal that evaluators define the success of CAE in terms of the quality of their working relationship with stakeholders. There were four major elements that emerged within this success dimension: leadership, communication, depth of participation, and the role of the evaluator in the evaluation.

There are many examples that demonstrate the critical role that stakeholder *leadership* plays in defining the relationship with evaluators, and this emerged as a significant aspect of success. Here is an example where program managers and leaders fostered positive relationships:

Program managers at each facility were partners in the evaluation, arranging interviews and focus groups at their sites and reporting on program milestones on a regular basis.

At the same time, we also found that stakeholder leadership could have enormous negative impact on the collaborative process and its perceived success. We observed that sometimes stakeholder leaders are motivated by political or nonrational concerns which run counter to the potential benefits of collaboration. The effects can be powerfully destructive, as this excerpt illustrates: "Senior leaders dictated the measures and, when the data did not support the desired outcomes, the project was abandoned."

Another important aspect of evaluator-stakeholder relations was *communication dynamics*, and our data revealed that continuous communication factored

into the evaluators' conception of success. But the flip side of communication was also evident. That is to say, when communication was poor or inconsistent, it could be highly detrimental to the collaborative process. Consider the following comment, associated with a project with only limited success:

Leaders decided what to do on the fly and often changed expectations for participants. When they did meet with an evaluator, it was for a short period of time and they often seemed distracted. [We] learned to meet with them during breaks and send late-night emails to one leader.

It is clear in our data that effective communication between stakeholders and evaluators is essential to the success of CAE. Such projects require an ongoing cycle of questioning and critical reflection relative to the learning generated throughout the process. This kind of effective communication facilitates the depth of participation of various stakeholders. Our data reveal that adherence to communication, partnership, and teamwork contributed to a much more active role for stakeholders in different aspects of the evaluation process and factored positively into the conception of success. A key to the success of the project was the organic nature of the process. In the words of one respondent, "All the various teams consulted with each other often as issues arose." Our data also show that, to a large extent, the success of the collaboration depends on the relevance of the evaluation and on stakeholders' ownership.

We also found that it is important to have the evaluator's and the stakeholders' roles identified and made clear from the beginning so that everyone understands what their involvement entails. More specifically, our results suggest that clarifying *the evaluator's role* beforehand is critical to setting expectations and avoiding misunderstanding and conflict later on.

We also looked at what evaluators considered to be specific reasons for CAE project success or the lack thereof in association with the "evaluator-stakeholder relations" success dimension. Of relatively high occurrence were considerations about the depth of participation and about evaluator-stakeholder relations. Ultimately, the evaluator-stakeholder relationship should be a two-way street, and to achieve the desired goals it is necessary to have commitment on all sides.

Stakeholder Interrelations

Workplace relationships have unique characteristics with important implications for the individuals in those relationships and for their work and productivity. These interrelations play a critical role in the development and maintenance of trust and positive feelings in any work environment. We found that stakeholder interrelations factored into evaluators' conceptions of success in CAE to a significant extent. There were three specific aspects that emerged: the level of agreement among stakeholders, stakeholder commitment, and support from program management.

Our data revealed that the CAE processes led to the alignments of stakeholders' interests and expectations that resulted in *agreement* among them, which factored into the evaluators' conception of success. For example:

The conflicts among the stakeholders were resolved early in the process. Communications among all the parties were clear and frequent during the evaluation.

On the other hand, lack of agreement among stakeholders could be a potent barrier to success:

There was a lot of tension about “who said what, to whom” and therefore disagreement with evaluation conclusions, and also whether things should even be written about within the context of the evaluation report as information sharing when the project funder was viewed as high risk for the tribes.

To a large extent, our data also showed that CAE both generated and benefited from mutual *commitment* from stakeholders to the evaluations. Making and keeping commitments are recognized by the evaluators as one of the most important aspects of stakeholder interrelations. We found many specific examples where evaluators identified stakeholders’ commitment as evidence of success and others where lack of commitment was a severe impediment to success.

To a large degree, our findings revealed that evaluators define the success of CAE in terms of a specific antecedent factor, namely, *supportive program management*. Many specific examples indicate that program management can influence CAE success either positively or negatively. According to one participant, “The fact that the management team is collaborative and cohesive has made our job easier.” It is important to note that management could also act as a barrier impacting negatively on evaluation success, as was the case for this participant: “The program developer and manager were not trustworthy, did not follow through with responsibilities.”

Finally, we looked at factors influencing the “stakeholder interrelations” success dimension and found the most frequently identified factor is the *relevance* of the evaluation, which is associated with identifying evaluation objectives and enhancing understanding of the program. As one evaluator mentioned: “The evaluation gained relevance and salience for stakeholders through participation.” *Stakeholders’ ownership* is also identified as a significant factor. There are many other factors associated with this success dimension; they mostly relate to interpersonal relations among stakeholders and their involvement in the evaluation process.

Stakeholder Engagement

In CAE projects, the decision is not whether to engage stakeholders or not, but when and how to successfully engage them. To a significant degree, our findings reveal that evaluators defined the success of CAE in terms of the meaningful engagement of stakeholders in CAE processes. Engagement is an iterative process that occurs throughout the evaluation process, beginning with consideration and scoping of key evaluation issues. Our data indicated that stakeholder engagement has the possibility of securing a wide range of benefits for the evaluation. Four main aspects emerged under this success dimension: dialogue, relevance, process use, and buy-in.

Generally speaking, it is important that good stakeholder interrelations are in place at the outset but, as our findings reveal, CEA can contribute to the development such relationships through *dialogue* and growth by engagement with the evaluation process. We also learned from evaluators describing projects that were less than successful that poor relations among stakeholders could be interpreted as constraints on dialogue and explaining the observed lack of success.

Evaluators provided many examples that show an increase in the evaluation *relevance* as a result of stakeholders' engagement in different evaluation processes including decision making. In the words of one evaluator, "the community member stakeholders were able to 'push back' when the evaluator and funder offered an approach that they found meaningless. Without the collaborative approach we don't think that would have happened." Stakeholder engagement implies a willingness to discuss issues of interest to stakeholders and, critically, to be prepared to consider making changes to the evaluation as a result of stakeholder engagement.

We also observed that embedding stakeholder engagement throughout the CAE projects has the possibility of increasing *process use* and, ultimately, the use of findings as is suggested by the quotation set out below:

The evaluation uncovered all the holes in the program and, since the stakeholders were integral parts of the evaluation, they felt that they were uncovering the problems in their design, not that an outsider was telling them something was wrong.

If CAE is about maximizing positive impact and changes, then stakeholder engagement is key because it is evident that it increases *stakeholder buy-in* in the evaluation. Evaluators' experiences show that stakeholder buy-in is a process of involving stakeholders in various aspects of CAE projects, including the decision-making process, in hopes of reaching a broader consensus and understanding. When stakeholders are not engaged meaningfully, the success of the evaluation might be jeopardized.

The top factor influencing the "stakeholder engagement" success dimension is the *relevance* of the evaluation, which in this dimension means that the priority is given to satisfying stakeholders' needs and interests in being engaged in CAE projects. For example, "the stakeholders were involved early in the process and their issues and questions were included in the evaluation." *Stakeholders' ownership* is also identified as a significant factor, as was explained by one evaluator: "All stakeholders had a common goal and were committed to the evaluation. Because of the stakeholder commitment, results were used as an opportunity to learn and grow."

Alignment of Evaluation Purpose

Developing a common understanding among key stakeholders of the purpose and objectives of the evaluation and the means and processes of accomplishing those objectives is very critical and has a significant impact on the success of CAE projects. Our findings reveal that evaluators defined the success of CAE in terms of the alignment of evaluation purpose with program community information needs. There were three major elements that emerged in association with this

success dimension: common understanding of the project objectives, clarity of the goals and processes, and the nature of stakeholder information needs.

There are many examples that demonstrate the importance of stakeholders' *common understanding* of the CAE objectives and how it factored in the evaluators' conception of success. In the words of one evaluator, "setting forth clear objectives greatly facilitated the design and conduct of this evaluation." A lack of common understanding runs the risk of increasing the likelihood of projects not achieving success, as demonstrated by the responses below. Note that sometimes this lack of consensus may derive from nonrational, political forces at play.

Developing a common understanding among the key stakeholders requires *clarity* in the description and communication of the objectives, processes, and expectations of the CAE projects. This is critical to ensure the CAE purpose alignment. The following quotations illustrate the point in the context of projects that were not successful:

The project was a process evaluation of a regional planning project funded through a federal grant. Although the grant specified what was needed for project implementation, it did not give guidance to the evaluator. So the purpose for the evaluation was not clear.

The program administrators were uncertain about what they wanted from the evaluation.

These examples demonstrate that clarity is critical to the planning and execution of CAE projects because it helps define the project scope. Related, stakeholders' information needs bring focus to CAE projects and are critical for all stages, including prioritizing resources and planning activities. In some cases we found that stakeholders' information needs were related to learning, improvement, and capacity building. On the other hand, evaluators pointed out that CAE projects that primarily focused on meeting the funding exigencies and/or accountability demands were significantly constrained in their success. As one evaluator explained, "[The] purpose of the evaluation was primarily to report to funders; evaluators had hoped for greater use of program evaluation for continued program design and improvement." It seems likely that most, if not all CAE projects would be at least partly accountability-oriented. Yet it seems plausible that when the agenda is more about learning and development, the likelihood of CAE success is increased.

When we looked at what evaluators considered to be specific *reasons for CAE project success* or the lack thereof in association with the "alignment of evaluation purpose" success dimension, we found that *information needs* was the most frequently identified factor. Also of relatively high occurrence were considerations about the *relevance* of evaluation and *depth of stakeholder* participation. Alignment of evaluation purpose is a process that implicates a direct role for stakeholders throughout the evaluation. Misalignment can lead to confusion, waste of resources, demotivation of both stakeholders and evaluators and, ultimately, severely limited CAE project success.

Resources

To a great extent, our data indicate that the available resources (i.e., time, money, and expertise) have great implications for the success of CAE projects. The amount of resources available could even influence a CAE project's rigour or the certainty of its findings, which, of course, factored into the evaluator's perceptions of success. There were three specific aspects that emerged: sufficient budget, time, and expertise.

Clearly, a sufficient *evaluation budget* is critical for effectively carrying out the evaluation processes and activities. We learned from evaluators describing successful CAE projects that these projects were adequately funded, which allowed for greater access or reach to program participants and increased sophistication in the processes. In the words of one participant:

The amount of funding for the evaluation was sufficient to allow for a rigorous quasi-experimental design with multilevel modelling. The technical sophistication of the evaluation made the results more credible.

Not surprisingly, evaluators describing projects that were less than successful commented that the inadequacy of the budget limited the involvement of stakeholders and compromised the quality of the data. In the example below, the respondent seemed to frame stakeholder participation as a potential cost-saving strategy.

There was a clear expectation early on for a high level of stakeholder involvement, but this was translated into a bare-bones budget based on the belief that the evaluators wouldn't have to put in as many hours if they weren't doing the actual data collection.

CAE can be expensive, particularly when a good number of program community members are involved in the joint evaluation work. Clearly, a sufficient budget is invaluable for ensuring the success of CAE projects.

Our data also reveal that an equally important aspect of the "resources" success dimension is the amount of *time* available for CAE projects. In many responses, the evaluators referred to the time to determine the success of the CAE. In some cases, this implicated the extent to which evaluators were unencumbered by other demands. In some cases, the availability of sufficient time allowed for increased relationship building and development between the evaluator and the stakeholders, which of course had a great influence on the success of the evaluation. On the other hand, evaluators describing less-than-successful CAE projects explained that the lack of time negatively influenced many aspects of the evaluations.

In addition, some evaluators reflected on the fact that human resources in terms of their own skills and *expertise* are factors of major importance to CAE success. In the words of one participant:

There are many players who have expertise in their role in the evaluation. The evaluators play a key role in the development of the evaluation. They have spent a great deal of time on background research and have improved the evaluation over time.

On the flip side, evaluators also communicated that a lack of expertise in evaluation processes could negatively influence the success of CAE. We observed the detrimental effects of involving program community participants for the wrong reasons (e.g., anticipated cost savings). Consider the following illustration:

The client wanted to be more involved in data collection to save costs. They thought their staff could collect the data rather than trained evaluators. The staff were unable or unwilling to collect data consistently or as accurately as trained evaluators.

Finally, we looked at what evaluators considered to be specific reasons for CAE project success or the lack thereof in association with the “resources” success dimension. Of relatively high occurrence were considerations about the *depth of participation* of the stakeholders in the process. This illustrates the significant linkage between the availability of resources (i.e., time and money) and the feasibility of having stakeholders meaningfully and deeply involved in the collaborative process. As one evaluator mentioned, “the stakeholders who were invested in the results were invested in the process.”

Other Success Dimensions

Two final dimensions of success of CAE emerged through our data analysis but were relatively infrequently identified. Still, they deserve mention here. First, *evaluation timeliness* was found to be important in a few instances. Timeliness is generally taken to mean adherence to decision-making cycles and processes. In the following example, we can see that the CAE mapped nicely onto the cyclic nature of the program:

Each cycle of the program was evaluated and the findings and recommendations taken on board in the next cycle of training. The constant improvement of the program was therefore also evaluated and celebrated. The evaluation, in these circumstances, meant that the success of the program grew and so did the motivation of the stakeholders and program participants.

We can see some interesting dynamic connections to the use of evaluation findings here. In other contexts, funding or staff turnover or other evolving program changes negated the potential for CAE success. In one case, the evaluation became obsolete.

A final dimension of successful CAE worth considering was *benefits to the evaluator*. The bulk of our foregoing analysis show that most conceptions of success implicated the program community alone (use of findings, process use, stakeholder engagement, purpose alignment), although evaluators were directly implicated in evaluator-stakeholder relationships. In the present case, the evaluator alone was the locus of benefit. Here is one illustration:

It was such a great project. I replicated it once with that same client, we co-presented the approach at a national conference, and I have just signed a contract to replicate in a second state.

We can see that pride and job satisfaction factored into evaluator-related benefits. In the example, dissemination and ongoing project opportunities defined the benefits. Another evaluator mentioned having published a paper from the CAE experience.

DISCUSSION

In this study we identify dimensions that help define successful CAE based on a secondary data analysis of the responses of a large group of evaluators who were asked to reflect on their experiences with highly successful and less-successful-than-hoped CAE projects. Seven key dimensions emerged from the data: use of findings, process use, stakeholder interrelations, evaluator-stakeholder relations, stakeholder engagement, alignment of evaluation purpose, and resources. Two other dimensions—evaluator benefits and timeliness—surfaced as well, but with much less frequency. The overlap in the findings (e.g., relevance, relationships, clarity of purpose, and communication) can be viewed as a reflection of their importance. Based on this, we propose a conceptual framework, as is set out below in [Figure 1](#), which captures not only these dimensions but also their relative prominence as derived from our sample of evaluators who implement CAE. That is, the magnitude of each of the shapes in the conceptual framework represents the low, moderate, or high factoring into the conception of success. The progression from left to right refers to temporal ordering (antecedent, process, and outcome). The progression from top to bottom refers to the locus of the observed factor: (a) evaluator, (b) program community, or (c) the interactive processes between them. We hope that this conceptual framework can be useful in terms of informing members of the evaluation community about the meaning of success, particularly those interested in developing tools and methods for measuring or operationalizing success in CAE.

Many of these dimensions and the relationships among them are not new, as can be seen in the literature, but it is important to note that they are empirically derived. The prominence of use—both the use of evaluation findings and process use—indicates the significance of such evaluation consequences in the field. Perhaps most notable is the reiteration of “soft” indicators, such as communication, and the importance of relationships as being central to the success of an evaluation. Others emphasize that the dynamic nature of success, depending on one’s point of view, may be thought of differently. But, as [Patton \(2010\)](#) suggests, we need to be aware of the complexity of the interplay among the contextual factors, stakeholder issues, and evaluation-management practicalities that shape the success, or the lack thereof, of an evaluation process. Framing our thinking as “incomplete success” rather than in binary terms (i.e., success/failure) is likely to be more helpful in the process of teaching us how to evolve and adapt.

Our study has limitations that should be mentioned. The findings are derived from a secondary analysis of data and therefore are indirect and highly inferential. In addition, the data are solely from the evaluators’ perspective, and thus represent one point of view and only one point in time. In this regard, we have observed that

Figure 1. Emergent framework for CAE success

our respondents tended not to take a critical stance toward their own approach, skills, or behaviour when describing success or especially lack of success.

Implications for Further Research

This work opens up some interesting possibilities for further research. Exploring the meaning of success from the perspective of stakeholders is one possible line of inquiry, in particular looking at inter-stakeholder and inter-evaluator/stakeholder differences and similarities in perceptions. As we have noted, the process is dynamic and changes over time, and so longitudinal designs would be required to capture the changes over a period of time. Another possible research direction could be to design a study with a specific and direct focus on examining success. The issue of ethics in CAE is another fruitful area of inquiry. Further, researchers may well ask from a realist perspective, “What are the contextual considerations and mechanisms that lead to success in different CAE contexts?”

Our study provided an especially rich resource for thinking about what success means in CAE, and we look forward to continuing the dialogue around these questions. We hope that this discussion will stimulate further empirical inquiry involving not only practitioners, but also stakeholders themselves.

ACKNOWLEDGEMENTS

A version of this article was presented at the annual meeting of the Canadian Evaluation Society, June 2016, St. John's, NL. The authors are grateful for financial support from the University of Ottawa. The opinions expressed in this article do not necessarily reflect those of the sponsor.

NOTES

- 1 A comprehensive technical report on the project (Shulha et al., 2015) is available at the following link: https://crecs.uottawa.ca/sites/crecs.uottawa.ca/files/shulha_et_al_2015.pdf
- 2 Note that a more detailed treatment of methods and results is available in the version of the paper presented at CES 2016, available in the CES grey literature data base available at <http://evaluationcanada.ca/>

REFERENCES

- Adams, A. E., Nnawulezi, N. A., & Vandenberg, L. (2015). ‘Expectations to change’ (E2C): A participatory method for facilitating stakeholder engagement in evaluation findings. *American Journal of Evaluation*, 36(2), 243–255. <http://dx.doi.org/10.1177/1098214014553787>
- Birkmayer, J. D., & Weiss, C. H. (2000). Theory based evaluation in practice: What do we learn? *Evaluation Review*, 24(4), 407–431. <http://dx.doi.org/10.1177/0193841X0002400404>

- Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35(1), 26–44. <http://dx.doi.org/10.1177/1098214013503699>
- Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not*. Oakland, CA: Berrett-Loehler.
- Connor, R. (2010). Reflections over 25 years: Evaluation then, now and into the future. *Canadian Journal of Program Evaluation*, 25(3), 125–136.
- Cousins, J.B. (1995). Assessing program needs using participatory evaluation: A comparison of high and marginal success cases. In J. B. Cousins & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 55–71). London, UK: Falmer.
- Cousins, J. B., Whitmore, E., & Shulha, L. M. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22. <http://dx.doi.org/10.1177/1098214012464037>
- Davies, H. T. O., Nutley, S., & Smith, P. C. (2000). *What works? Evidence based policy and practice in public services*. Bristol, UK: Policy Press. <http://dx.doi.org/10.1332/policy-press/9781861341914.001.0001>
- Elbaz-Haddad, M., & Savaya, R. (2011). Effectiveness of a psychosocial intervention model for persons with chronic psychiatric disorders in long-term hospitalization. *Evaluation Review*, 35(4), 379–398. <http://dx.doi.org/10.1177/0193841X11406080>
- Fetterman, D. F., Kaftarian, S. J., & Wandersman, A. (Eds.). (2015). *Empowerment evaluation* (2nd ed.). Los Angeles, CA: Sage.
- Gervais, M., & Joubert, P. (2010). (Eds.). As I Recall—Or How to Take Advantage of Less-Than-Successful Evaluation Experiences, *Canadian Journal of Program Evaluation*, 25(3), xi–xxvii.
- Granger, R. C., & Maynard, R. (2015). Unlocking the potential of the ‘What works’ approach to policymaking and practice: Improving impact evaluations. *American Journal of Evaluation*, 36(4), 558–569. <http://dx.doi.org/10.1177/1098214015594420>
- Hawkins, P. (2010). Successful evaluation management: Engaging mind and spirit. *Canadian Journal of Program Evaluation*, 25(3), 27–36.
- Liket, K. C., Rey-Garcia, M., & Maas, K. E. H. (2014). Why aren't evaluations working and what to do about it: A framework for negotiating meaningful evaluations in nonprofits. *American Journal of Evaluation*, 35(2), 171–188. <http://dx.doi.org/10.1177/1098214013517736>
- Marek, L. I., Brock, D. P., & Savla, J. (2015). Evaluating collaboration for effectiveness: Conceptualizations and measurement. *American Journal of Evaluation*, 36(1), 67–85. <http://dx.doi.org/10.1177/1098214014531068>
- Mills, H., Crone, D., James, D. V. B., & Johnston, L. H. (2012). Exploring the perceptions of success in an exercise referral scheme: A mixed method investigation. *Evaluation Review*, 36(6), 407–429. <http://dx.doi.org/10.1177/0193841X12474452>
- Moehr, J. R., Schaafsma, J., Anglin, C., Pantazi, S. V., Grimm, N. A., & Anglin, S. (2006). Success factors for telehealth: A case study. *International Journal of Medical Informatics*, 75(10–11), 755–763. <http://dx.doi.org/10.1016/j.ijmedinf.2005.11.001>

- Owen, J. (2010). Successful or not: It depends on your frame of reference. *Canadian Journal of Program Evaluation*, 25(3), 85–93.
- Patton, M. Q. (1978). *Utilization-focused evaluation* (1st ed.). Newbury Park, CA: Sage.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2010). Incomplete successes. *Canadian Journal of Program Evaluation*, 25(3), 151–163.
- Perrin, B. (2010). Learning from evaluation misadventures: The importance of good communication. *Canadian Journal of Program Evaluation*, 25(3), 115–124.
- Poulin, M., Harris, P. W., & Jones, P. R. (2000). The significance of definitions of success in program evaluation. *Evaluation Review*, 24(5), 516–536.
- Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, 35(1), 116–119. <http://dx.doi.org/10.1177/1098214013499439>
- Raphael, S., & Stoll, M. A. (2006). Evaluating the effectiveness of the Massachusetts workforce development system using No-shows as a nonexperimental comparison group. *Evaluation Review*, 30(4), 379–429. <http://dx.doi.org/10.1177/0193841X05282268>
- Sawhill, J. C., & Williamson, D. (2001). Mission impossible? Measuring success in nonprofit organizations. *Nonprofit Management & Leadership*, 11(3), 371–386. <http://dx.doi.org/10.1002/nml.11309>
- Scriven, M. (2016). Roadblocks to recognition and revolution. *American Journal of Evaluation*, 37(1), 27–44. <http://dx.doi.org/10.1177/1098214015617847>
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & Al Hudib, H. (2015). Evidence-based principles to guide collaborative approaches to evaluation: Technical report. Ottawa, ON: Centre for Research on Educational and Community Services. https://crecs.uottawa.ca/sites/crecs.uottawa.ca/files/shulha_et_al_2015.pdf
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & Al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 193–215. <http://dx.doi.org/10.1177/1098214015615230>
- Sturges, K. M. (2015). Complicity revisited: Balancing stakeholder input and roles in evaluation use. *American Journal of Evaluation*, 36(4), 461–469. <http://dx.doi.org/10.1177/1098214015583329>
- Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present and future. *American Journal of Evaluation*, 22(3), 387–395. <http://dx.doi.org/10.1177/109821400102200316>
- Walter, A. J., & Scholz, R. W. (2007). Critical success conditions of collaborative methods: A comparative evaluation of transport planning projects. *Transportation*, 34(2), 195–212. <http://dx.doi.org/10.1007/s11116-006-9000-0>
- Wandersman, A. (2009). Four keys to success (theory, implementation, evaluation and resource/system support): High hopes and challenges in participation. *American Journal of Community Psychology*, 43(1-2), 3–21. <http://dx.doi.org/10.1007/s10464-008-9212-x>

Wimbush, E. (2014). Reshaping evaluation to enhance utilization in Scotland: The role of intermediary bodies in knowledge-to-action strategies. In M. L. Loud & J. Mayne (Eds.), *Enhancing evaluation use: Insights from internal evaluation units* (pp. 83–112). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781506335124.n5>

AUTHOR INFORMATION

Elizabeth Whitmore is Professor Emerita, School of Social Work, Carleton University. Her interests include collaborative approaches to evaluation, gender issues in evaluation, and social justice.

Hind Al Hudib is a Doctoral Candidate, Faculty of Education, University of Ottawa. Her interests include evaluation policy and organizational capacity for evaluation, and collaborative approaches to evaluation.

J. Bradley Cousins is Professor Emeritus and Senior Researcher, Centre for Research on Educational and Community Services, University of Ottawa. His interests include collaborative approaches to evaluation, evaluation use, and evaluation capacity building.

Nathalie Gilbert is a Doctoral Candidate, Faculty of Education, University of Ottawa. Her interests include collaborative approaches to evaluation, evaluation in health systems, and patient engagement.

Lyn M. Shulha is Professor Emerita, Faculty of Education, Queen's University. Her interests include evaluation use, collaborative approaches to evaluation, and student assessment.