

Reflections on the meaning of success in collaborative approaches to evaluation: Results of an empirical study.

Elizabeth Whitmore*, Hind Al Hudib**, J. Bradley Cousins**,
Lyn Shulha***, & Nathalie Gilbert**

DRAFT

Paper presented at the annual meeting of the Canadian Evaluation Association, St. John's, Nfld. 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. A version of this paper has been submitted for publication consideration.

June 2016

*Carleton University, ** University of Ottawa, ***Queen's University

Reflections on the meanings of success in collaborative approaches to evaluation: Results of an empirical study.

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 et al., 2013; Shulha et al., 2016) that developed a set of evidence-based principles to guide evaluation practice in contexts where evaluation knowledge is collaboratively produced by evaluators and stakeholders. In this paper, we examine data from 320 evaluators' responses to our (2012) survey that explored questions about their practice experience with CAE. The survey consisted of two sections: one probing a CAE project that the respondent considered 'highly successful' and one probing another CAE project thought to be 'far less successful than hoped'. In addition to quantitative ratings, in each section we asked respondents to "tell us a story (in 500 words) about the reasons why this was a highly successful (or less than successful) CAE." The results revealed that, while evaluators do in fact define success according to consequences such as use of evaluation findings and process use, they are more likely to think in terms of evaluator-stakeholder and stakeholder-stakeholder relationships and in terms of the extent to which evaluation purposes aligned with stakeholder information needs. Based on the findings, we propose a conceptual framework as an aid to thinking about success in CAE and we comment on an agenda for further inquiry.

INTRODUCTION

Initially in evaluation discourse, program success was generally equated with having achieved the program goals. Scriven (2016) begins a recent commentary on evaluation's role in this respect with a classic dictionary definition of program evaluation: "Summative evaluation is defined as '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... (p.28). In articulating this perspective, the author 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, it follows that evaluation success has been achieved – in terms of intended use by intended users (Patton, 1978, 1997). Others, using words like ‘effectiveness’ (Elbaz-Haddad & Savaya, 2011; Liket, 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, 2014; Liket et al., 2014; Sturges, 2015). Fetterman, Kaftarian and Wandersman (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 models that have emerged over the many years of dialogue and development of empowerment evaluation – notably the 3 and 10 step models – “enhance the probability of program success” (p.9).

One 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* entitled “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 volume include communication, the importance of standards and evaluator competencies, and engagement with key stakeholders.

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 Connor adds that “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” (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). In the case examples describing their experiences with CAE discussed in this article, program advocates considered an evaluation to be ‘unsuccessful’ because it did not support their views, while the evaluator, attending to the integrity of the process, felt that the evaluation was successful in that it adhered to professional standards. 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” (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 with 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 Savia (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 (2013) 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 provided 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 Scholtz (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 non-profits (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.

Research Objectives

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 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 et al., 2016). A comprehensive technical report on the project (Shulha et al., 2015) is available at the following link: [insert URL here]. 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.

Data structure

In the main phase of our data collection, 320 practising evaluators completed our online instrument which included quantitative and qualitative items. Participants identified a CAE project from their own experience that they considered to be highly successful, and they provided a set of responses about that project including two open-ended questions: (i) "What were the top 3 reasons why this collaborative approach to evaluation was highly successful?" and (ii) "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 paper.

It should be noted that each evaluator participant 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 item 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.

We also looked, secondarily, at corresponding data from the first open-ended question that detailed evaluators' opinions about factors influencing success (or lack of it). In looking at these data, we relied on codes that had been previously developed and applied for the purposes of generating principles to guide CAE practice. Details about the coding structure and data quality assurance for that project are elaborated by Shulha et al. (2015).

Coding structure and data quality assurance

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.

Insert Table 1 about here

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 applied 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 our 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. In doing so, we will also present corresponding data associated with reasons for success, coded in the previous phase of our project.

Two analysts (Al Hudib and Cousins) assumed principal responsibility for coding and analysing the data using 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 used these responses to conduct an inter-coder agreement analysis on two occasions. Table 2 reveals the results for this analysis. It should be noted that we decided to drop one code (conformity to CAE models or approaches) because it was essentially irrelevant, and to create a new emergent code (stakeholder engagement) which turned out to be fairly prevalent. Having jointly decided this, we reviewed previously coded data and reapplied the codes accordingly.

Insert Table 2 about here.

As shown in Table 2, our analysis was based on only a limited amount of independently coded data and that coding accuracy improved over occasions. We were confident after the final occasion that we had a clear understanding of code meaning and appropriate

application. As mentioned, all data were reviewed after each occasion to update/adjust the coding based on our deliberations.

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 ordered by evaluator participant (rows), and included all data (columns) provided in response to the open-ended questions about (i) ‘other details’ highlighting specific segments of text associated with the respective code and (ii) ‘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 are some illustrations.

Evaluation was integral to program planning, development, and improvement. Evaluation results were used by program management to make changes to improve the program.

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.

These evaluation results fueled policy change at the state level.

It is clear, in these examples that CAE feeds into an improvement process more so than accountability interests. Part and parcel of program improvement is **‘learning.’** Our data also reveal that the conceptual use of evaluation findings was evident. The following verbatim quotation illustrates the point.

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, here with CAE serving 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 evaluator spoke in terms of symbolic use of findings and its disingenuous qualities. In the words of some participants,

Senior leaders dictated the measures and when the data did not support the desired outcomes, the project was abandoned.

Evaluation was state-required, which the superintendent resented.

The CEO was more interested in impact data that would prove the organization's value to the members than he was in building capacity for ongoing learning into the organization.

We also learned from evaluators describing projects that were less than successful that the **non-use** of evaluation data could be taken to define observed lack of success. In their words,

At the end of the evaluation period, it appeared that curricular changes (moving toward a problem-based curriculum) would strengthen the deliberate use of technology to "do things differently and do different things." However, that initiative faltered

The approach resulted in evaluation findings that the technology worked, but the decision was made based on fiscal, rather than evaluation, findings. The trust of the stakeholders was undermined as a result.

We can see that there is a political and/or non-rational element to this discourse.

Insert Figure 1 about here

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. Figure 1 shows the array of reasons that were given. Of relatively high occurrence were considerations about the *relevance* of the evaluation to the information needs of the stakeholders. One evaluator, for example, mentioned that the evaluation was "focused to the more critical areas," while another suggested that the "outcomes reflected the inputs of all the participants so the results were accepted."

Other evaluators identified an organizational *culture* that is evaluation-friendly as being an important reason for success. For example,

Program managers respected front-line staff, involved them in the evaluation process, and showed them that they valued their input.

Front line staff, who were responsible for collecting the data, did not understand the importance of getting it collected accurately

This latter quotation, in reference to a less-than-successful CAE project, shows how an organizational culture that does not fully embrace evaluation can lead to negative outcomes. Another reason that was given, and that often related to lack of success in CAE, had to do with the explicit *purposes* of the evaluation. Clarity of purpose appears to be paramount, as shown in the next quotation.

Ambivalence ...about purpose and target of the evaluation (is it about learning/improvement or proving to our funders that we are worth re-funding)? This resulted in confusion and lack of buy-in throughout client system.

Another important factor associated with the 'use of findings' success dimension had to do with *interpersonal relationships* among stakeholders. 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."

Figure 1 shows that there was a variety of other factors or reasons for success or lack of it that were associated with this success dimension. Many of these had to do with relationships and personal qualities of stakeholders and/or evaluators. Stakeholder engagement with, and ownership of, the evaluation appeared to factor in as well.

Process Use

Process use is a complex phenomenon associated with benefits arising from stakeholder proximity to the evaluation. This phenomenon was found to factor in to evaluator's conceptions of success in CAE to a significant degree. There were three specific aspects which 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. To follow are some illustrations:

[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.

The center's director, a physician, eventually took a program evaluation short course from me and learned to appreciate the value of creating her own logic models. Subsequently, she received a local grant to take a healthy school menu into charter schools with a built in evaluation. She calls herself an evaluator, too. For me, this is a success story demonstrating my skills in using collaborative evaluation approaches (practical rather than theoretical) to improve health programs that directly affect the lives of individuals and programs in the community.

[The stakeholders] also learned about evaluation and they use the evaluation instruments and any other skills to run their program (learning). With the help of the stakeholders, I hired local people to collect data for the evaluation; so these skills stay in their communities and they use it all the time.

We can see 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 **use of findings** but our data reveal a **linkage** between the two. Specifically, process use is mostly about learning but as we can see in the following examples such learning may be interconnected with and even augment conceptual use or learning from the findings of the evaluation.

Also impactful because youth influenced what we explored and, as it turned out, the areas of greatest surprise to the adults were on topics the youth indicated should be explored and put in measures on.

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.

Stakeholders came to meetings ready to discuss solutions and were not defensive about the problems they uncovered.

Post-project interviews with stakeholders indicated that adults paid more attention to the findings than prior needs assessments in part because of the level of collaboration with youth from around the state.

Some of these responses shared by evaluators 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. In the following examples transformation takes the form of the development of relationships, organizational and program structures, and understandings about program and organizational capacity.

Primary Health Care in Aboriginal Community - changes to self-determination and control, increasing community capacity in primary care.

[The stakeholders] became invested and more comfortable in sharing authority and participating with each stage of the evaluation. Ownership was high throughout the process. I know that their expectations for future evaluations (and evaluators) are different - they will expect and seek out collaborative (done with) and reject didactic (done to) approaches.

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.

Insert Figure 2 about here

Finally, as we can see in Figure 2, we looked at which were the main **factors** influencing the process use success dimension. The figure reveals 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 quotations:

Cultural appropriate evaluation and the instruments developed with the help of the stakeholders.

Participating in evaluation meetings allowed stakeholders to identify problems with their program

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. As one participant put it “clear learning purpose for the evaluation; emphasis was on process not on evaluating individual performance.” The second had to do with stakeholder *information needs* and their specific interests in the evaluation and the findings it was likely to generate, as the following excerpts attest: “Stakeholders were interested in the outcome;” “authentic interest in outcomes of the evaluation.” In addition to these factors, Figure 2 reveals an array of others, many reflecting the importance of relationships among stakeholders and evaluators as well as communication processes.

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 also 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. Examples include:

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

There was agreement between stakeholder interest groups, the program administrators on the purpose of program evaluation and its role in program improvement. There was stability between staff and the program director who had good working relationships with providers, program participants, advocates, and staff.

Expectations and unexpressed assumptions were more likely to be discussed and agreement reached throughout the evaluation process.

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.

Promises and plans were made, only to find that the representatives' bosses disagreed or forced approval through a lengthy process.

The purpose is to develop an online game to teach scientific thinking. A game group and a science education group not [see] things [the] same [way]. [They had] different goals.

To a large extent, our data also revealed 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. Some illustrations are set out below.

Unique collaboration across systems (education, public health, law) that was "ripe" with partners committed to a shared goal/purpose. That commitment allowed sticky questions and issues to be addressed without having ego get in the way.

The stakeholders helped in the decision making and they felt that the evaluation was their project. They felt that they fully participated, their opinion was heard and they used the results of the evaluation to improve their program.

Three existing regional groups gather monthly to identify goals that have an impact on children in their community and what they might do as group members to facilitate the process of meeting those goals.

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

We can see that such commitment is vital in driving stakeholders toward mutually desirable goals. Without commitment, there is no common purpose. It was clear lack of commitment could be taken to indicate the observed lack of success. As some comments revealed:

Only a few stakeholder representatives were able to make firm commitments regarding the evaluation.

The originator of the program lost interest in this program half way through.

The superintendent did not want the evaluation, but it was required by the state. Some of the school staff were afraid to talk to the evaluators for fear that their comments would identify them (confidentiality was promised, and comments were aggregated and reworded to hide identities). The

superintendent did not "like" any form of bilingual education and so informed the evaluators that even if we identified this as a need, and regardless of what the research showed about the model we might suggest, she would not consider it.

While more than commitment alone is needed for success, through CAE, stakeholders were in a better position to commit on an informed commitment, without which, projects would be at risk of failure.

To a large degree, our findings reveal 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. For example:

The fact that the management team is collaborative and cohesive has made our job easier.

The sponsor saw the value of including the evaluand, specifically the managers of the program, because the evaluation team saw that the program would benefit greatly from the evaluation while also conducting a neutral, somewhat rigorous evaluation.

The executive director was and continues to be very committed to evaluation and listening to their stakeholders, and the active use of evaluation findings.

But management could be also act as a barrier impacting negatively on success. In evaluators' words:

The program developer and manager were not trustworthy, did not follow through with responsibilities.

The stakeholders were not seen as important as much as those to be controlled. The punitive dynamics of the posturing in the program led to challenges and resistance. Weak leadership coupled with political agendas nixed this improvement process before it really got started.

The primary problems often involve management-related issues.

Very motivated mid management and implementers but they had to convince top management again, again and again...

Our impression from these findings is that supportive program management factors significantly into evaluator's conceptions of success.

Finally, Figure 3 presents a set of the principal factors influencing the 'stakeholder interrelations' success dimension. The most frequent factor is the *relevance* of the evaluation, which is associated with identifying evaluation objectives and enhancing understanding of the program. As one evaluator mentioned: "Collaborating ensured a high level of understanding about the program and the key issues to be investigated". Another evaluator similarly said: "Participation increased everyone's stake in the evaluation results and how those results were achieved. The evaluation gained relevance and salience for stakeholders through participation." *Stakeholders' ownership* is also identified as a significant factor:

“Collaboration facilitated stakeholder buy-in with the evaluation design and implementation, which in turn smoothed the pathway for the evaluation happening”. Another evaluator also said: “For me the key to whether there is good collaboration is the degree to which individuals in the project care about it”. As can be seen in Figure 3, 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.

Insert Figure 3 about here

Evaluator-Stakeholder Relations

To a significant 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 are some examples 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.

The Program Director and her staff have been highly supportive of the evaluation work. They have established an internal evaluation team that communicates and collaborates with us on a frequent basis. The internal team conducts the process evaluation while we focus on the outcome evaluation.

A new administrator had been stymied in attempts to move the stakeholders of this program to do a self-study, so they brought me in. Overall the experience was positive and the outcomes well received. 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.

At the same time, we also found that stakeholder leadership could have enormous negative impact on the collaborative process and its perceived success. In the following examples taken from reflections on less than successful CAE projects, we can see that sometimes stakeholder leaders are motivated by political or non-rational concerns which run counter to the potential benefits of collaboration. The effects can be powerfully destructive.

Senior leaders dictated the measures and, when the data did not support the desired outcomes, the project was abandoned.

After two and a half years, the leaders and managers have found out how important it is to the funder that the evaluation produces data. None of them wants to participate in deciding key indicators of

performance. My team has been denied access to data and, at the same time, criticized for not producing reports they want or interpretations of the data. My philosophy is collaborative and they have had little tolerance for my requests for their input. It has been a thankless job.

The end of year report was 'wordsmithed' in two six-hour sessions by the funder before it was accepted and we were paid. So in this case having the participation of the funder and program implementers who wanted to only see the positive was not collaborative but coercive.

The President would prioritize and design the research (he is a researcher). While he would acquiesce and include some collaborative elements in the design, he was not interested in anything beyond academic outcomes, so did not support collaboration. I was new and did not know that he would not support. I spent a lot of time getting to know the front-line developers, who were really excited about being involved. Ultimately, though, their input or ideas were excluded.

Another important aspect of evaluator-stakeholder relations was **communication dynamics** and our data revealed that continuous communication factored in the evaluators' conception of success. According to one evaluator:

Continuing dialogue kept open the lines of communication and problems that arose were solved within days. This was a collaborative effort and resulted in positive program outcomes and an evaluation process that went smoothly.

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 these comments, associated with projects with only limited success.

Communications between a primary funder and the collective evaluation team (evaluators and participating stakeholders) were difficult. Program funders and managers made some strategic changes in program planning, and did not share this information with program staff or the evaluation team or participating beneficiaries.

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 from these examples 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. Set out below are some illustrations.

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, especially for me as the external evaluator. There was a clear role and area of focus for the external evaluator that allowed me to be involved as a witness and facilitator of the process without slipping into a role of service provider to district-level stakeholders.

A key to the success of the project was the organic nature of the process. All the various teams consulted with each other often as issues arose.

Our data also reveal that, to a large extent, the success of the collaboration depends on the relevance of the evaluation and on stakeholders' ownership. In the words of some participants,

Those involved had a stake in making the initiative work. They believed in it and wanted to know how to improve or make it work better. There was time to engage with participants (evaluation funded for multiple years). For me the key to whether there is good collaboration is the degree to which individuals in the project care about it. Also if the evaluation is designed to allow collaboration through rather than a more narrow focus on collecting so data, analyzing it and telling us if the program is working.

A "participatory evaluation" was required of the Commission overseeing and funding this evaluation. We, as evaluators, defined what "participatory" meant and recruited stakeholders to participate in all aspects of the evaluation. This evaluation study was the first of its kind. It was highly successful in that stakeholder participation was strong, and stakeholders endorsed the approach.

Our data also showed 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 show that clarifying **the evaluator's role** beforehand is critical to setting expectations and avoiding misunderstanding and conflict later on. Some illustrations are provided below.

We described it as collaborative and a partnership from the beginning. It took a little while to earn their trust, but when we did it was phenomenal. So much learning in both directions!

We simply make it clear that they are not paying us to make the evaluation problems go away – that we are not going to do evaluation of them or for them. We are going to do this together for the best result.

Stakeholders/funders perceived me, and sought me out as the content expert and were ready to turn the evaluation entirely over to me. However, I brought a collaborative approach to the table from the outset and they LOVED it. They became invested and more comfortable in sharing authority and participating with each stage of the evaluation. Ownership was high throughout the process.

The evaluation team established procedures for data collection and closely monitored how projects collected and submitted data. Any data-collection issues were quickly resolved.

We co-created the logic model, approach, tools, sampling strategy. Stakeholders alone collected all the data. We co-analyzed the data and interpreted the results. I as evaluator modeled how they might describe their data and gave them sample bar charts, polar graphs and other tabular ways to present their data. Stakeholders alone reported back the findings to their communities. Awesome! It was such a great project!

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. Figure 4 shows the array of reasons that were given. Of relatively high occurrence were considerations about the depth of participation and about evaluator-stakeholder relations.

For example, one evaluator mentioned that “The evaluation team visited program sites periodically not only to collect data but to also answer questions and build relationships”, while another said that “Little or no ego was involved. Evaluators and stakeholders were all at the table for the same reason: to help”. As a result, as one participant mentioned that “The evaluator became trusted to represent stakeholder interests up the line”. Ultimately, the evaluator-stakeholder relationship should be a two-way street and in order to achieve the desired goals it is necessary to have commitment on all sides.

Insert Figure 4 about here

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 define 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 revealed 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 through engagement with the evaluation process. The following examples illustrate:

Collaboration means that stakeholders are talking with one another and often are learning things about an initiative prior to any evaluation reporting. Stakeholders are therefore primed for the evaluation findings.

Collaboration builds and strengthens relationships that are important for the development and sustainability of an initiative, including the provision of support for changes to an initiative.

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. In their words,

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

Less cohesion amongst stakeholders

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. The following example from less than successful evaluation confirms this point:

The client process delegated all decisions to the evaluators and simply nodded at results. They were not willing to engage or take responsibility for contacting potential participants from their list to participate in baseline.

Our data also revealed 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 quotations 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.

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.

If CAE is about maximising positive impact and changes, then stakeholder engagement is a key because it is evident that it increases **stakeholder buy-in** in the evaluation, as the following quotations demonstrate.

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

The stakeholders helped in the decision making and they felt that the evaluation was their project; they felt that they fully participated, their opinion was heard and they used the results of the evaluation to improve their program.

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, as we can sense from the following evaluator's comment.

It was a data analysis and writing exercise. Stakeholders were mandated to report. The data was already provided and was not clean. Regional program differences made it difficult to tell a cohesive performance story, and every stakeholder wanted their program profiled uniquely.

Finally, Figure 5 presents a set of the principal factors influencing the 'stakeholder engagement' success dimension. The top factor 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. As one evaluator put it: "All parties involved were highly engaged in this evaluation work and contributed extensively to the content of instruments... the final products really reflected the main elements of the program". Another evaluator similarly said: "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". As we can see in Figure 5, there are many other factors associated with this success dimension that are mostly related the participation of stakeholders in the evaluation process.

Insert Figure 5 about here

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 non-rational, political forces at play.

The purpose of the project was to evaluate the programs/areas offered by a community college. The definition of what constituted a "program/area" ranged from a single course taught in one content area to academic degree or technical certification programs. Content stakeholders often saw limited benefit to the process or the products produced. Administrative stakeholders reviewed the results of the evaluations often years after the data was collected and analyzed, which impacted the relevance of any decisions made based on the evaluation findings and recommendations.

Disagreement among stakeholders about what the evaluation should be measuring and how.

Hidden agendas on the part of all parties (including evaluators) that we were unable to surface through the process.

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 since it helps define the project scope. Also, our findings show that common understanding and clarity of purpose are also critical when it comes to various stakeholders information needs. Information needs bring focus to CAE projects and is critical for all stages including prioritizing resources and planning activities. Our data reveal that in some successful CAE projects, stakeholders' **information needs** were related to learning, improvement, and capacity building. Some examples are set out below.

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

Funders wanted to know if the project worked. Implementers liked the project and thought it was working but did not have any data to show it was. Intended project beneficiaries liked the project and wanted to see it continue.

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. In their words,

The accountability focus was a bottleneck, because none of the stakeholders would 'share' the unintended consequences, the challenges or shortcomings.

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.

Stakeholders participated in order to receive funding for the evaluation and to demonstrate eager cooperation with the funders (feared budget cutbacks or losing competition during anticipated amalgamations).

The focus was to prepare a report on a national program and to ensure regions were profiled. The exercise was clearly intended to ensure the program 'looked good'. It was a data analysis and writing exercise.

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 as shown in Figure 6. The Also, of relatively high occurrence were considerations about the *relevance* of evaluation and *depth of stakeholder* participation. For example, one evaluator mentioned that "The evaluation design was constantly reviewed, revised and discussed during regular meetings with the client"; another stated that "The final products really reflected the main elements of the program". When the purpose was not clear, "stakeholders had difficulty seeing the relevance or value of the evaluation". 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, de-motivation of both stakeholders and evaluators and, ultimately, severely limited CAE project success.

Insert Figure 6 about here

Resources

To a great extent, our data reveal 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 their words,

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

There were sufficient resources so that every person with a disability who wanted to provide input could.

Small evaluation project but it provided sufficient resources to ensure a good sample of respondents.

The Program Director has provided the budget necessary to successfully conduct both the internal and external evaluation components.

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. The verbatim quotations set out below illustrate this point. The first one 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.

The evaluation was greatly hampered by an unrealistic budget for evaluation.

Additional collaboration would have been helpful, but very expensive in view of the distances and other travel considerations.

This last comment suggests that CAE is can be expensive particularly when a good number of program community members are involved in the joint evaluation work. Clearly, these responses shared by evaluators reveal that 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. Here are some illustrations.

Team members dedicated time to developing the evaluation.

I was on leave from the university. My income was secure. My professional ideals could dominate. I extended leave three months in order to do the job because of the long front-end work both to become trusted and to make sense of the native-American perspective on the materials and instruction and match it to a literacy model and evaluation model. Time. I was full time and internal.

It was possible to have several meetings of the entire evaluation team with the organization workers to discuss and develop the evaluation questions. This allowed us to build rapport, trust, caring relationships and respect with one another horizontally.

We can see that 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. Consider the following:

I think mainly that I didn't have enough time for front-end work and for constantly being around to work out the complexity of this minimally funded project on both program and evaluation.

I only wish time/money had allowed for greater participation.

The co-evaluators' time was, however, not freed up, and they continued to manage their own areas by phone, while evaluating a peer area.

The timeline was too short to take all differences into account, so some key stakeholders/collaborators stopped participating.

As we can see, these comments add nuance and emphasis to the importance of having sufficient time and money to successfully handle CAE processes. 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 their words,

[The evaluation] was sponsored by the state oversight department, which also provided a skilled data technician who synchronized offender and recidivism data from two different state-level sources. This was a demanding process but, once it was over, everyone had confidence in the data.

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. In the following example, we can see the detrimental effects of involving program community participants for the wrong reasons (e.g, anticipated cost savings).

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. Figure 7 shows a set of reasons that were provided. 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". Another evaluator explained that "because stakeholders participated in regular meetings to review interim findings, they were able to put the results to immediate use".

Insert Figure 7 about here

Other Success Dimensions

Two final dimensions of success of CAE emerged through our data analysis but were relatively infrequently identified. Still, we are of the view that 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.

The stakeholders had changed from the time of the planning to the time of the evaluation and many were not aware of the purpose. The project had finished so the learnings had no impact on any change to the project.

Essentially, the evaluation, in this case, 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 are some illustrations:

I feel fortunate to be the internal evaluator for an agency that was one of the sites. The culture of constant learning and welcoming the evaluator into all phases of the project from implementation to wrap up is refreshing

It was a 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 latter 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 reflections of a large group of evaluators; reflections of highly positive and less-successful-than-hoped CAE evaluation experiences. Seven key dimensions emerge 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 8, that 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: (i) evaluator, (ii) program community or the (iii) 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.

Insert Figure 8 about here

Many of these dimensions and 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 our respondents tended not to take a critical stance towards 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.

Our data have 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 not only by practitioners, but perhaps also by stakeholders themselves.

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.
- Birckmayer, J.D. & Weiss, C.H. (2000). Theory based evaluation in practice: What do we learn? *Evaluation Review*, 24(4), 407-431.
- Brinkerhoff, R.O. (2003). *The success case method: Find out quickly what's working and what's not*. Oakland, CA: Berrett-Loehler Publishers.
- Connor, R. (2010). Reflections over 25 years: Evaluation then, now and into the future. *Canadian Journal of Program Evaluation*, special issue, 125-136.
- 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.
- 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: Falmer.
- Davies, H.T.O., Nutley, S., & Smith, P.C. (2000). *What works? Evidence based policy and practice in public services*. Bristol, UK: Policy press.
- 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.
- Fetterman, D.F., Kaftarian, S.J. & Wandersman, A. (Eds.) *Empowerment evaluation* (Second Edition). Los Angeles: Sage.
- Gervais, M. & Joubert, P. (2010). (Eds). As I Recall—Or How to Take Advantage of Less-Than-Successful Evaluation Experiences, *Canadian Journal Program Evaluation*, special issue.
- 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.
- Hawkins, P. (2010). Successful evaluation management: Engaging mind and spirit. *Canadian Journal of Program Evaluation*, special issue, 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.
- Marek, L.I., Brock, D.P. & Savia, J. (2015). Evaluating collaboration for effectiveness: Conceptualization and measurement. *American Journal of Evaluation*, 36(1), 67-85.
- Mills, H., Crone, D., James, D.V.B., & Johnston, L.H. (2013). Exploring the perceptions of success in an exercise referral scheme: a mixed method investigation. *Evaluation Review*, 36(6), 407-429.
- 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.
- Owen, J. (2010). Successful or not: It depends on your frame of reference. *Canadian Journal of Program Evaluation*, special issue, 85-93.

- Patton, M.Q. (1978). *Utilization-focused evaluation* (First edition). Newbury Park, CA: Sage.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (Third edition). Thousand Oaks, CA: Sage.
- Patton, M.Q. (2010). Incomplete successes. *Canadian Journal of Program Evaluation*, special issue, 164-178.
- Perrin, B. (2010). Learning from evaluation misadventures: The importance of good communication. *Canadian Journal of Program Evaluation*, special issue, 164-178.
- 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.
- 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.
- Sawhill, J.C. & Williamson, D. (2001). Mission impossible? Measuring success in nonprofit organizations. *Nonprofit Management and Leadership*, 11(3), 371-386.
- Scriven, M. (2016). Roadblocks to recognition and revolution. *American Journal of Evaluation*, 37(1), 27-44.
- 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*. doi:10.1177/1098214015615230
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & Al Hudib, H. (2016). Evidence-based principles to guide collaborative approaches to evaluation: Technical report. Ottawa: Centre for Research on Educational and Community Services.
https://crecs.uottawa.ca/sites/crecs.uottawa.ca/files/shulha_et_al_2015.pdf
- Sturges, K.M. (2015). Complicity revisited: Balancing stakeholder input and roles in evaluation use. *American Journal of Evaluation*, 36(4), 461-469.
- Torres, R.T. and Preskill, H. (2001). Evaluation and organizational learning: past, present and future. *American Journal of Evaluation*, 22(3), 387-395.
- 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, 3-21.
- Wandersman, A. (2015). Getting to outcomes: An empowerment evaluation approach for capacity building and accountability. In Fetterman, D.F., Kaftarian, S.J. & Wandersman, A. (Eds), *Empowerment Evaluation* (Second Edition). (pp.150-164). Los Angeles: Sage.
- 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.

Author Biographies

Elizabeth Whitmore is Professor Emerita, School of Social Work, Carleton University. Interests: collaborative approaches to evaluation; gender issues in evaluation; social justice.

Hind Al Hudib is Doctoral Candidate, Faculty of Education, University of Ottawa. Interests: Evaluation policy and organizational capacity for evaluation; collaborative approaches to evaluation.

J. Bradley Cousins is Professor Emeritus and Senior Researcher, Centre for Research on Educational and Community Services, University of Ottawa. Interests: Collaborative approaches to evaluation; evaluation use; evaluation capacity building.

Lyn M. Shulha is Professor Emerita, Faculty of Education, Queen's University. Interests: Evaluation use; collaborative approaches to evaluation; student assessment.

Nathalie Gilbert is Doctoral Candidate, Faculty of Education, University of Ottawa. Interests: Collaborative approaches to evaluation; evaluation in health systems; patient engagement.

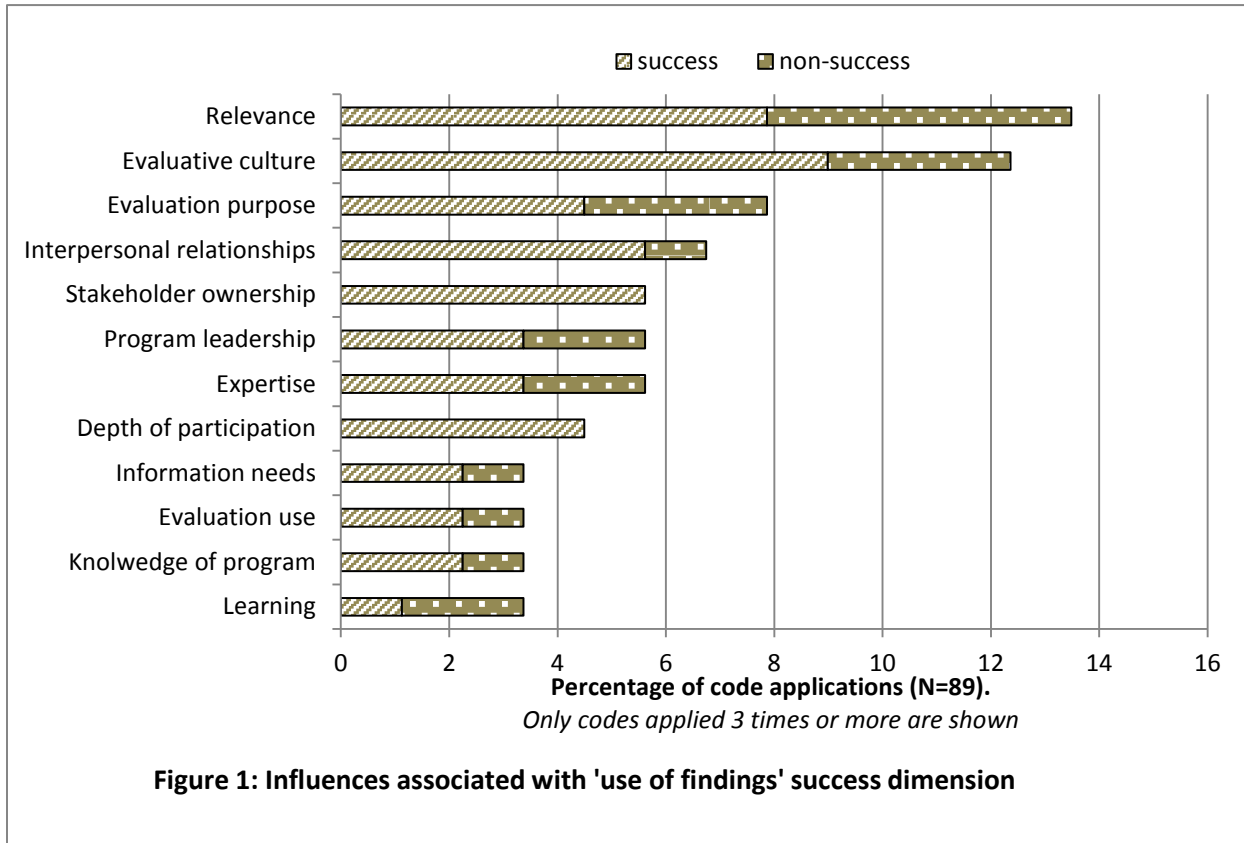
Table 1: Emergent Codes by Frequency of Application

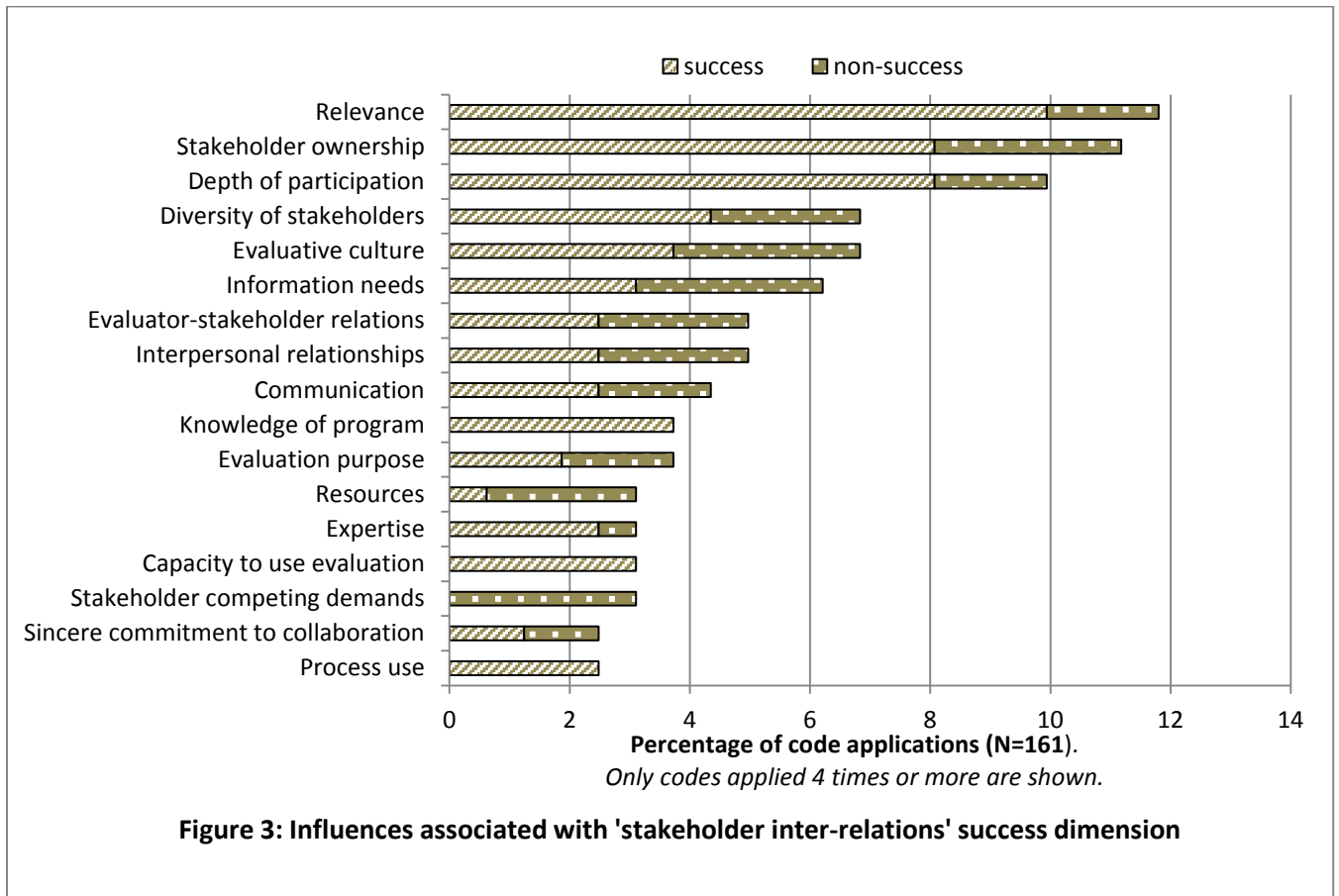
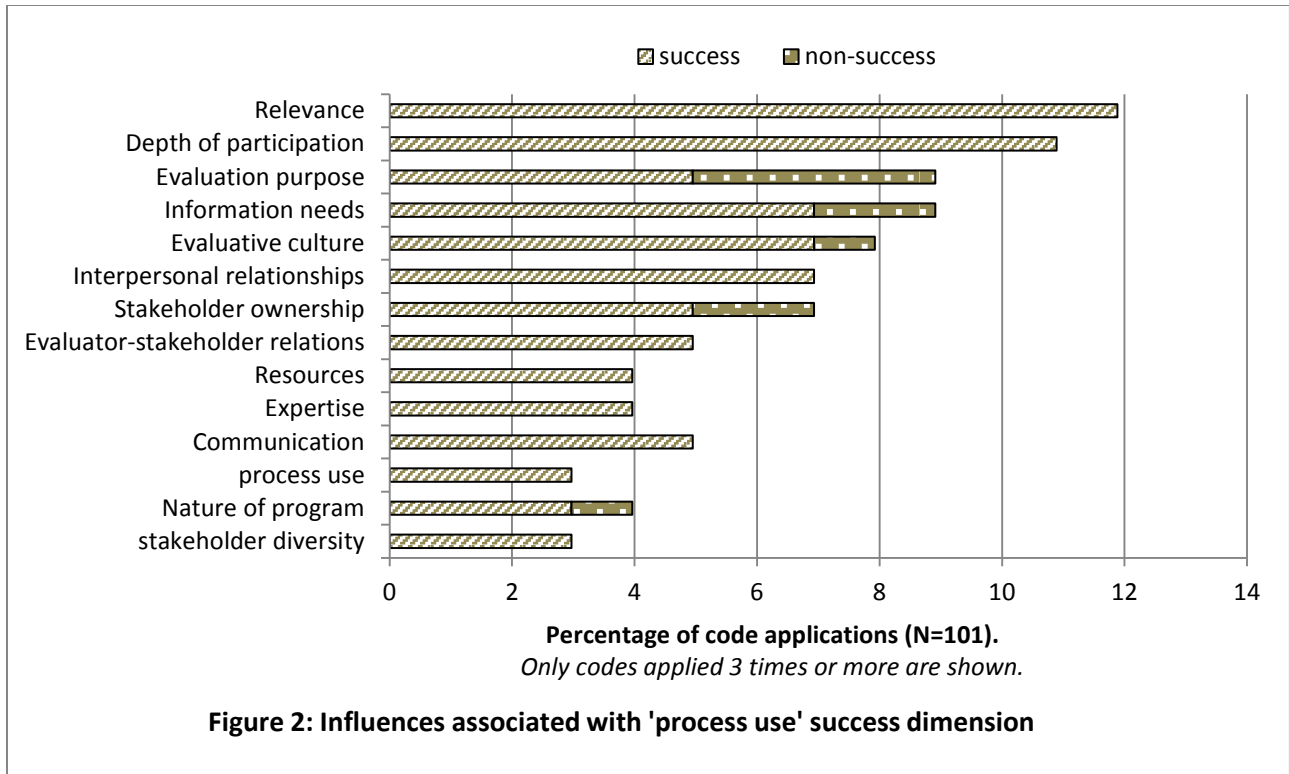
Code	Description	# Code Applications
Defining Success/Non Success		
Use-F	Use of findings or instrumental (program change), conceptual (learning about program) and symbolic benefits of the evaluation.	25 moderate
Use-P	Process use or benefits to stakeholders (individual, team, organization) of the evaluation 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

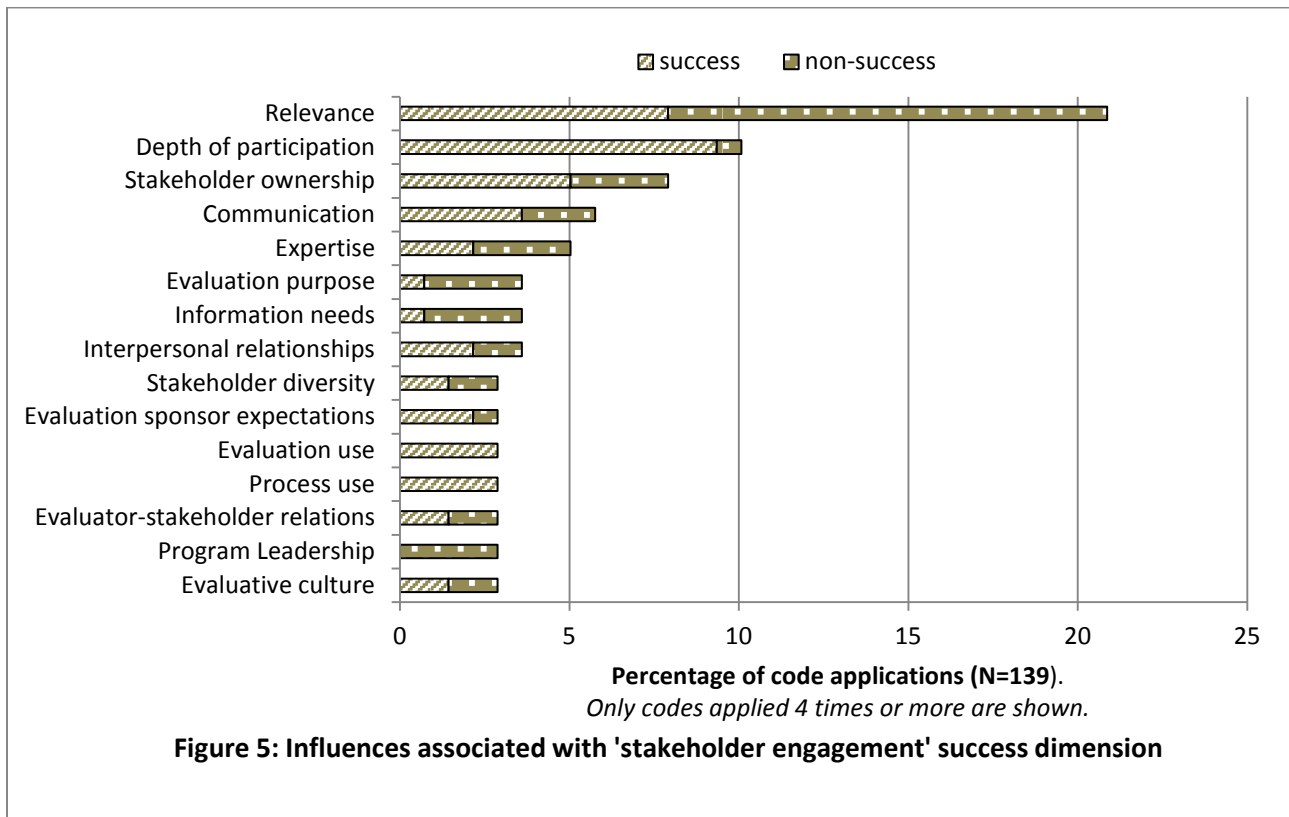
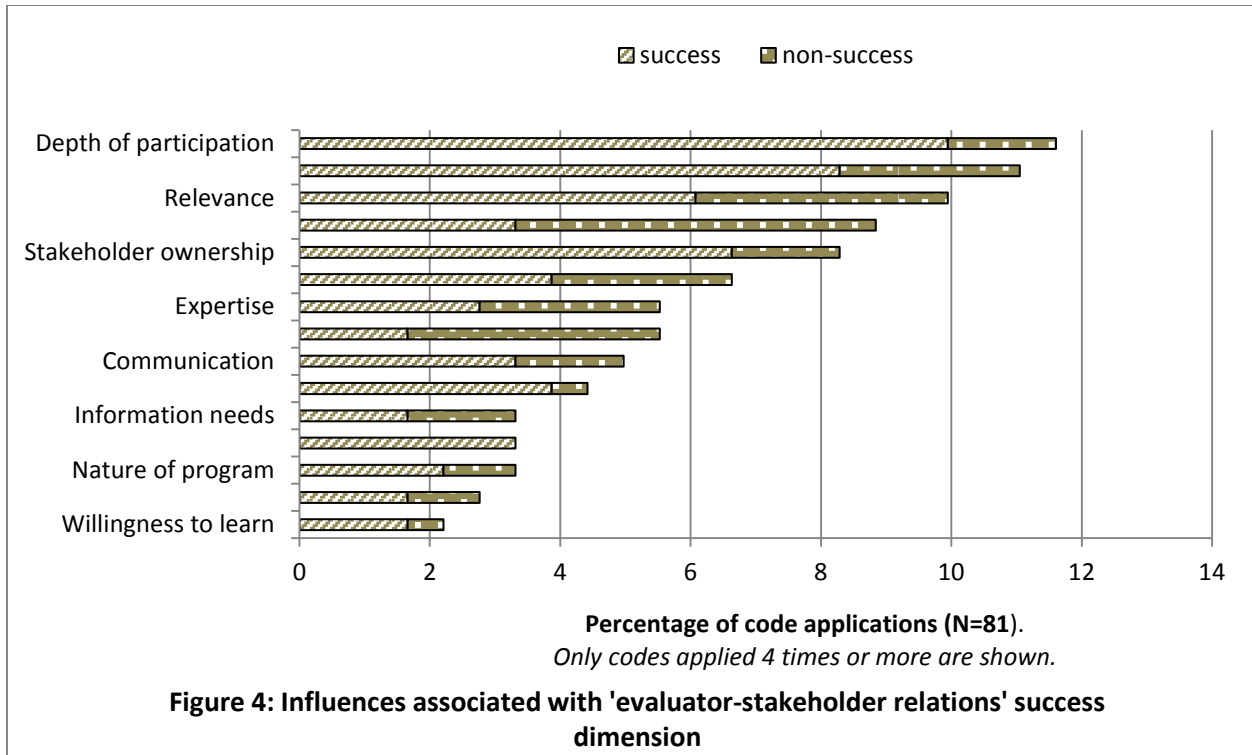
Table 2: Results of Inter-coder Agreement Analysis

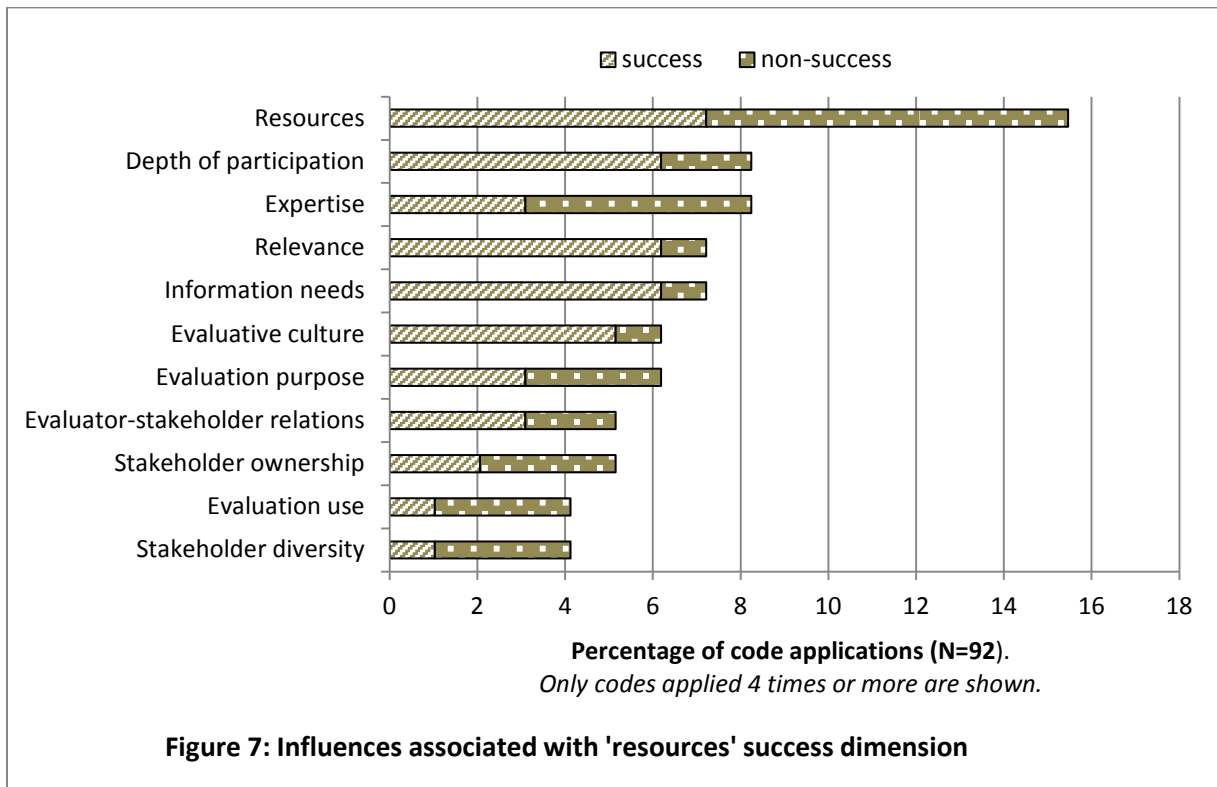
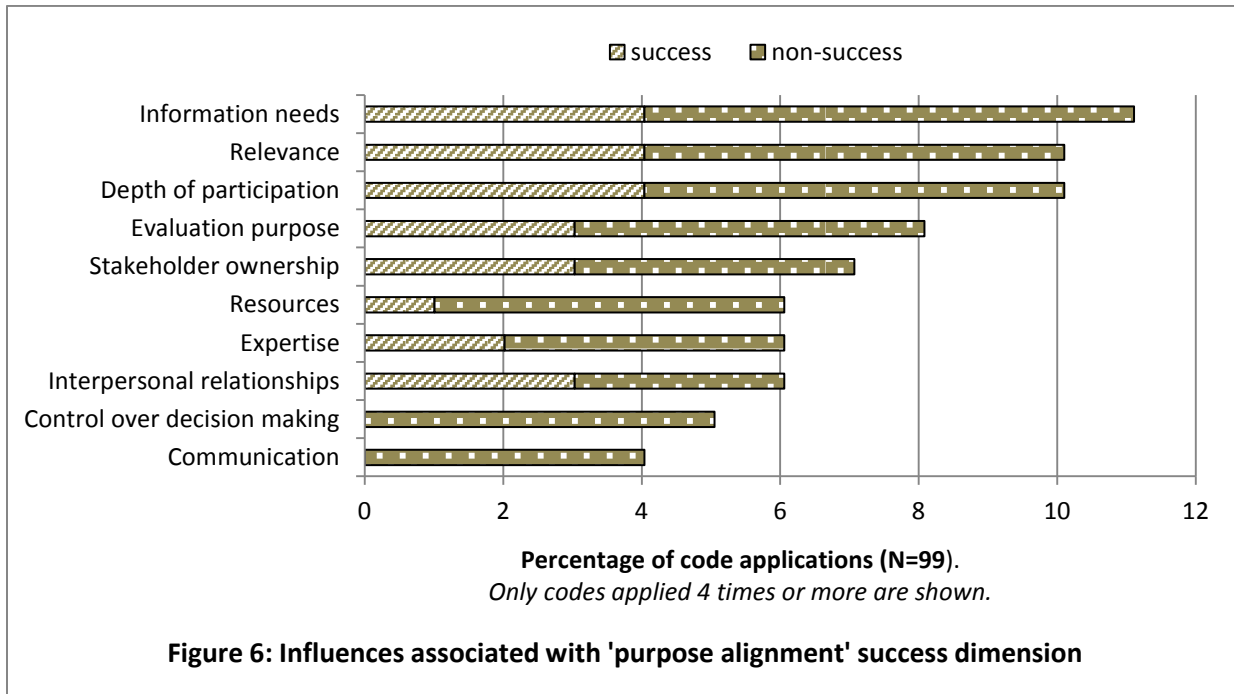
Independent coder application	Occasion 1*	Occasion 2*
Same category, same code	8 (47%)	10 (62%)
Same category, different code	6 (35%)	1 (6%)
Different category	3 (18%)	5 (31%)

*N=10 participant responses per occasion.









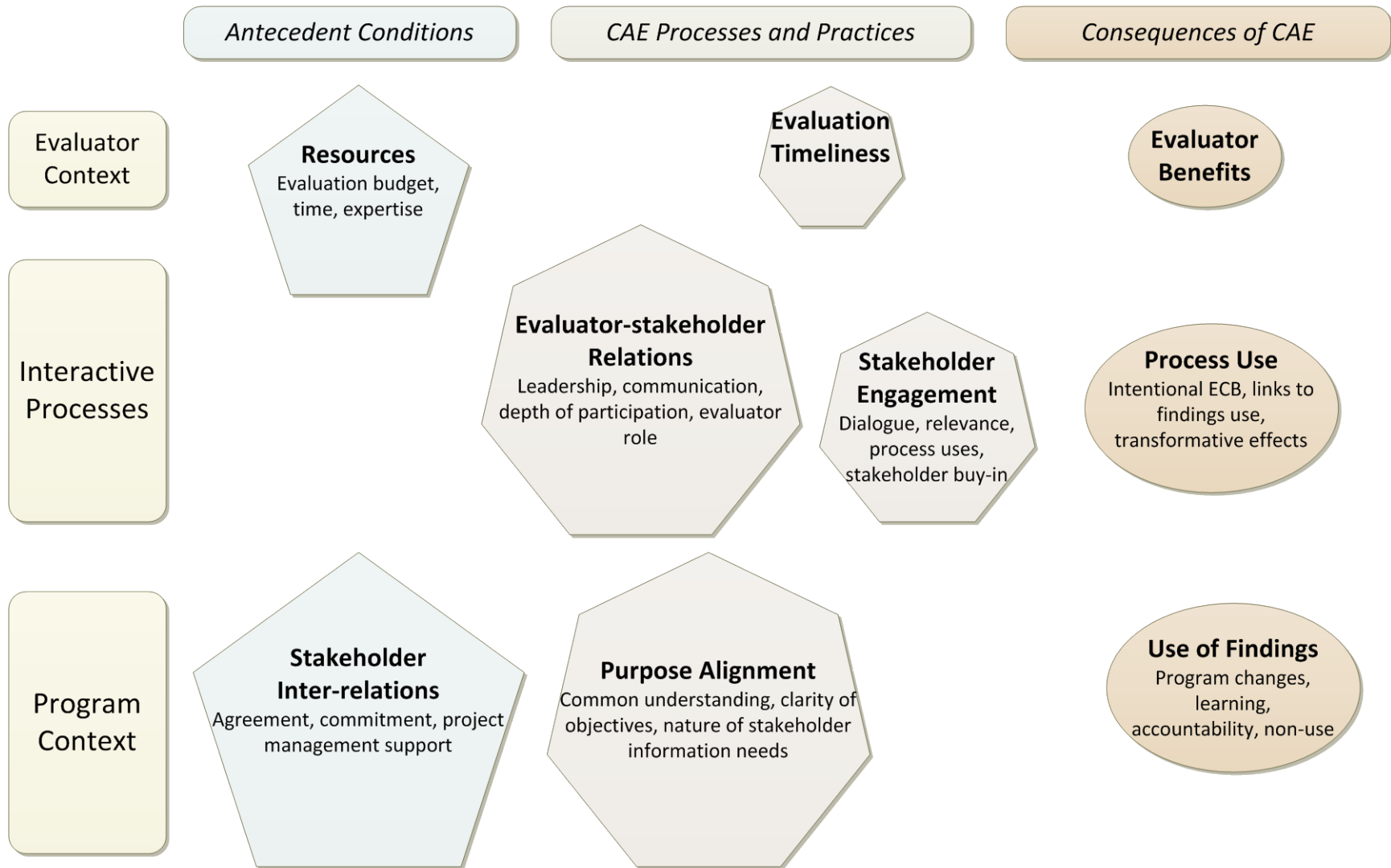


Figure 8: Emergent framework of CAE success