Understanding of evaluation capacity building in practice: a case study of a national medical education organization

Aimee J Sarti, Stephanie Sutherland, Angele Landriault, Kirk DesRosier, Susan Brien, Pierre Cardinal

Department of Critical Care, The Ottawa Hospital, The Royal College of Physicians and Surgeons of Canada, Ottawa, ON, Canada

Introduction: Evaluation capacity building (ECB) is a topic of great interest to many organizations as they face increasing demands for accountability and evidence-based practices. ECB is about building the knowledge, skills, and attitudes of organizational members, the sustainability of rigorous evaluative practices, and providing the resources and motivations to engage in ongoing evaluative work. There exists a solid foundation of theoretical research on ECB, however, understanding what ECB looks like in practice is relatively thin. Our purpose was to investigate what ECB looks like firsthand within a national medical educational organization.

Methods: The context for this study was the Acute Critical Events Simulation (ACES) organization in Canada, which has successfully evolved into a national educational program, driven by physicians. We conducted an exploratory qualitative study to better understand and describe ECB in practice. In doing so, interviews were conducted with program leaders and instructors so as to gain a richer understanding of evaluative processes and practices.

Results: A total of 21 individuals participated in the semistructured interviews. Themes from our qualitative data analysis included the following: evaluation knowledge, skills, and attitudes, use of evaluation findings, shared evaluation beliefs and commitment, evaluation frameworks and processes, and resources dedicated to evaluation.

Conclusion: The national ACES organization was a useful case study to explore ECB in practice. The ECB literature provided a solid foundation to understand the purpose and nuances of ECB. This study added to the paucity of studies focused on examining ECB in practice. The most important lesson learned was that the organization must have leadership who are intrinsically motivated to employ and use evaluation data to drive ongoing improvements within the organization. Leaders who are intrinsically motivated will employ risk taking when evaluation practices and processes may be somewhat unfamiliar. Creating and maintaining a culture of data use and ongoing inquiry have enabled national ACES to achieve a sustainable evaluation practice.

Keywords: evaluation capacity building, data use, program evaluation, quality improvement, leadership

Introduction

Evaluation capacity building (ECB) is a topic of great interest to many organizations as they face increasing demands for accountability and evidence-based practices. Evaluation capacity is about getting people in organizations to look at their practices and processes through a disciplined process of systematic inquiry. ECB is about helping people ask questions and then go seek out empirical answers. While various definitions of ECB exist, most writers in the domain of ECB agree that ECB is about building the knowledge, skills, and attitudes of organizational members; the sustainability of...
rigorous evaluative practices; and providing the resources and motivations to engage in ongoing evaluative work.\textsuperscript{1–6}

The fundamental premise underlying this definition is that the organization is committed to internalizing evaluation processes, systems, policies, and procedures that are evolving and self-sustaining. It also suggests that the organization must foster a learning culture that values trust, risk taking, openness, curiosity, inquiry, and champions the ongoing learning of its members.\textsuperscript{2,7} Based on our review of the ECB literature, we identified several key strategies that were consistently given priority among evaluation scholars. Furthermore, these strategies also resonated with our experiences and involvement in medical education organizations. Though some strategies were labeled differently than we may see in medical education and some were articulated in overlapping ways, the elements that guided our investigation included evaluation knowledge, skills and attitudes, and sustainable evaluation practice. Sustainable evaluation practice included subelements such as use of evaluation findings, shared evaluation beliefs and commitment, evaluation frameworks and processes, and resources dedicated to evaluation.\textsuperscript{1,2,6,8,9}

In an effort to contribute to the growing and evolving knowledge base about evaluation capacity within medicine, we undertook an exploratory case study with the national Acute Critical Events Simulation (ACES) program to better understand the extent that the organization utilizes ECB strategies. We also wanted to document lessons learned from their engagement in evaluation with the hope that other organizations will find their experiences useful as they progress toward sustainable evaluation practices.

Context
ACES is a national educational organization with the aim to ensure that health care providers (individuals and teams) from various clinical backgrounds become proficient in the early management of critically ill patients, given their individual scope of practice and the clinical milieu in which they work. Nurses, respiratory therapists, and physicians who are the first to respond to a patient in crisis come from various disciplines and practice in diverse milieus. Their experience managing acutely ill patients is often very limited given the low incidence of critical illness. Yet, clinical studies indicate that it is the early recognition and management of patients that is most effective in lowering both patient morbidity and mortality. For instance, randomized controlled trials and guidelines emphasize the importance of the “golden hour” in patients with conditions such as myocardial infarction, stroke, and sepsis.\textsuperscript{10–14} In the case of patients with septic shock, each 1-hour delay in the administration of antibiotics increases mortality by 7%.\textsuperscript{15}

ACES programmatic content includes various simulation modalities delivered online or face-to-face as well as books and didactic materials. It also includes instructor certification courses. Most of the educational material has been customized to meet the needs of different groups of learners. This organization was initially developed from the vision and efforts of a small collective of Canadian critical care physician leaders who volunteered their time and expertise. It has successfully evolved into a national education program, has more recently been acquired by the Royal College of Physicians and Surgeons of Canada (RCPSC), and continues to advance and grow. ACES also comprises the infrastructure, processes, and personnel required to develop, customize, and update the educational material, to train instructors and course coordinators, to deploy courses, and to monitor and evaluate the program itself. ACES includes various courses customized for nurses and respiratory therapists, physicians both in practice and in training, and from different specialties (eg, family medicine, internal medicine, and anesthesiology). However, the flagship course of ACES is the national ACES course offered to all critical care fellows in Canada. The course has been offered yearly for the last 13 years. This study focused on the national ACES training program. This study was exempt from approval by the Ottawa Hospital Research Ethics Board as it was part of a larger quality improvement initiative. All participants provided verbal consent.

Methods
This study was an extension of a larger comprehensive needs’ assessment undertaken as part of a quality improvement initiative for national ACES. We undertook a qualitative study to understand the extent of ECB within a particular medical educational organization. We opted for a more prestructured qualitative research design as we wanted to bound the study within a set of ECB strategies, yet at the same time, we needed to maintain enough flexibility to allow emergent findings.\textsuperscript{16} Thus, we conducted semistructured interviews with program leaders and instructors so as to gain a richer understanding of evaluative processes and practices from their personal narratives.

Qualitative data
Interviews were conducted between October 2012 and May 2014. Participant selection was carried out using a purposive sampling to identify individuals who would provide a balanced representation across a wide variety of characteristics.\textsuperscript{17}
Interviews were confidential, and no incentives were provided. For the comprehensive quality improvement initiative, participants included program instructors from different specialties, ACES faculty members, and health care professionals from different backgrounds with representation from each of the 10 provinces in Canada. A total of 15 interviews were conducted with these individuals to gain a broad and comprehensive understanding of the ACES program. The participant population included physicians (n=11), nurses (n=2), and respiratory therapists (n=2). In gathering our interview sample, we learned that the ACES organization has a relatively flat organizational structure in that many of the physician teachers in the program are also considered leaders. To further explore the ECB construct within this organization, we performed an additional six interviews with national ACES leaders and instructors. Due to the organizational structure, all six leaders turned out to be instructors as well. Overall, a total of 21 individuals participated in the semistructured interviews. Qualitative interview participant characteristics are displayed in Table 1.

Semistructured interview guides were designed to follow a broad, predetermined line of inquiry that was flexible and could evolve as data collection unfolded, permitting exploration of emerging themes. Interview guides were created by an interdisciplinary team of investigators with expertise in medical education, simulation, program evaluation, sociology, and qualitative research methods. Interviews were audio-recorded and transcribed verbatim (refer Supplementary materials for the interview guides). Qualitative data analysis of the comprehensive data set included the application of inductive coding techniques. The research team followed Creswell’s coding process where data are first explored to gain a general sense of the data and then coded. These codes were described and collapsed into themes. Thematic saturation of the data was achieved in that no new themes or subthemes were found in the data. The analysis team consisted of two researchers who participated in coding training and meetings to develop the qualitative codebook. During development of the codebook, any disagreements in coding were discussed until consensus was reached. In an iterative process, additional interviews were performed with national ACES leaders and instructors to further explore the elements ECB and to validate the specific ECB constructs. These interviews were also audiorecorded, transcribed, and coded by the analysis team, with the consistent application of inductive coding techniques. All qualitative data were entered into NVivo software for data management purposes.

**Results**

Themes from our qualitative data analysis included the following: evaluation knowledge, skills, and attitudes, use of evaluation findings, shared evaluation beliefs and commitment, evaluation frameworks and processes, and resources dedicated to evaluation. Representative quotes are provided to clarify/explain each of these aforementioned themes later.

**Evaluation knowledge, skills, and attitudes**

Our analysis revealed that a central component toward an ongoing drive for evaluation capacity came from the program leaders and instructors. In terms of reasons for conducting evaluation, one of the leaders emphasized that internal motivation was a driving factor:

> Well, I think it’s just an internal motivation. Just the fact that we want to get better. We want to improve. We want to be the best program in early resuscitation in the world… period.

So, that in effect forced us to think of an evaluation strategy. In terms of external motivation, well I think those were less important, honestly. There was really no one forcing us to introduce the evaluative component. Maybe the only external factor that played a role was academic recognition.

In terms of knowledge and skills regarding evaluation, our interviews with the leaders revealed that they were perhaps self-effacing when describing their evaluation approach. That is, there was clearly an evaluation structure in place and it was embedded in the program design from the outset; however, leaders acknowledged that they were only at the beginning

### Table 1 Characteristics of interview participants

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mountain</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>Prairies</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Ontario</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>Quebec</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Atlantic</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td><strong>Specialty/discipline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical care</td>
<td>16</td>
<td>76</td>
</tr>
<tr>
<td>Pediatric critical care</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Anesthesia</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Surgery</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Family medicine</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Nurses</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Respiratory therapists</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

*Notes:* aThe regions of Canada have been divided into the following way: Mountain includes British Columbia and Alberta, Prairies include Saskatchewan and Manitoba, and Atlantic includes all Atlantic Provinces. bIndividuals were classified under their current practice specialties. Note that an individual may be practicing in more than one specialty.
stages of evaluation work and they wanted to do more complex evaluation work in terms of strengthened processes:

We are evaluating a two-day course and we are evaluating it at a very low level, you know, “Did you like the course?” … low in the Kirkpatrick level. We were able to link it with some competencies but this is why we want to undertake a more systematic look at our data. We recognize that we have to do more of this [evaluation].

Despite leaders’ feelings of not doing enough evaluation work and wanting to do more comprehensive evaluation, national ACES had clearly delineated processes for implementing and sustaining the evaluation that was in place at the time of this study:

There are guidelines in how we collect and provide feedback to the instructors because the data that comes back is not necessarily de-identified so we need to put a filter there and not send the data directly back to the instructor. So, there are some guidelines but there could be more structure.

From the interviews with leaders and instructors, the motivation to do evaluation was driven by an intrinsic desire to continually make the program better. Though leadership stated that they “didn’t know much about evaluation”, the assumption was that evaluation was a good thing to do and added value to ongoing decision-making. The expectation was that the evaluative work would not only continue but also there was a desire to employ a more sophisticated program evaluation cycle. In terms of designing the evaluation efforts, one of the leaders explained the process this way:

We had an employee with a Master’s degree in education, so he designed the evaluation of the ACES course, and he would spend time putting the evaluation report together. This person has left but we’ve taken the same evaluation format and given this task to somebody else within our pool.

Interestingly, although the leaders were self-proclaimed “nonevaluators”, they were savvy enough to build in more of a scholarly or academic component of the evaluative approach:

I think you build in a scholarly approach to the design of the course. There are some questions that get asked of all students, some that get asked of all preceptors or instructors and/or course directors that are uniformly asked across every course. They are embedded in the material and they are done so by design with the evaluative component in the design. That peer evaluation process and the evaluation of the course is embedded into the design as opposed to the usual after thought.

The implementation process was described as “straight-forward” whereby the evaluative feedback as well as the post-session debrief material was aggregated and entered into the ACES database. Evaluation data would be discussed during annual steering committee meetings, though there were often many informal conversations regarding ongoing challenges and successes. Many of the participants noted that the biggest barrier in conducting more evaluation-related activities was a lack of time. One of the program leaders/instructors simply stated “we need more time to learn about more about program evaluation, especially how we can better utilize data currently collected as well as plan for ongoing data capture.”

Use of evaluation findings

ACES leadership embedded evaluation as an element in their annual program planning processes, which turned out to be a key component in creating a culture of data use with the organization.

I think you need to review this course yearly. If you are offering it every year you need to be aggressive about getting feedback from the participants and finding out what works and what doesn’t work, especially if it’s a new course. If there are little things like this scenario didn’t work or this equipment was terrible, but also if there are bigger things like you didn’t talk about this and I really needed to know that.

Once the annual evaluation report is out we go through it at our National Steering Committee meeting, so everybody will come to that meeting having read the report. The report usually says something like, “we’re doing very, very well in high fidelity simulation, the phone simulation has been very well received, however the simu case or the virtual simulation has not been received so well…why is that?” Then we explore and we propose change and we follow up on it the year after. The course can be dramatically changed as a result of the evaluation feedback.

In this organization, evaluation data act as both a trigger for programmatic changes and a means to celebrate program successes. Program leadership shared with us that it was important for the culture of the organization to share program successes widely and frequently.

Shared evaluation beliefs and commitment

Deep and meaningful emotional components connecting the leadership/instructors of national ACES were identified. Clinical instructors expressed positive emotions by being part
of the network, and such engagement appears to be strongly linked to this network of like-minded educators,

You feed off the energy of the others, and they feed off your energy… it is a fantastic success component. The initial energies that people had to make this program happen were just dynamic. Everyone had new ideas, everyone wanted to work together. There was one person who organized the initial think tank but he kind of tied the whole thing together… he put it into a form… like molding clay into a form but the clay consisted of a whole bunch of people with lots of ideas. That is what makes this such a winning team. I think any place discovers when they are in business, one of the key components has to be the involvement of people who love it and like to think outside the box because as soon as you do that, you have a recipe for success.

Being a part of this network is intellectually very pleasing, it’s emotionally rewarding. You end up feeling like you are part of this community. As a physician, you end up being a better physician because you are learning things from others that you can apply at the bedside. Taken those features and a love of teaching… these instructors are really teachers at heart… it becomes a chance to show their skills.

The affective dimension of the instructors/leaders of national ACES was evident throughout many of the interviews. We heard of deep and meaningful emotional ties that connected members of the organization. The organizational culture within national ACES serves to drive the passion to continually strive for success and this translates to an interest in using data for continual improvements.

Evaluation frameworks and processes

We learned that the structure of conducting evaluation was built into the organization at its concept. Similarly, program leaders and instructors had a strong desire to use evaluation results for ongoing programmatic improvement. Interestingly, most of the discussions about evaluation use occurred in an informal manner. This is likely due to the strong social network ties among the leaders/instructors.

So post course we always have an informal faculty debrief, it typically happens just before the end of the course or immediately after the course so all the faculty are still there. We have a quick debrief about what went well, concerns, do we need changes, and it’s not just about content but about processes and logistics and things like that…

With the informal meeting we all (instructors) get together for dinner. I’m not so sure that we learn so much that is new about the course but the one piece of feedback we do get immediately is the emotional response of the residents. Like if a simulation was tough… those are the types of things we discuss. Like for this one resident the simulation was particularly traumatic. You know, she cried or he was very stressed so there’s some debriefing that needs to go on immediately with the residents and the instructors will bring that information to our meeting.

Perhaps the small size of the staff working with the instructors/leaders contributes to the organization-wide understanding of the importance of evaluation processes. As evidenced by the data, many of the programmatic discussions and decisions occur informally. Another program leader/instructor noted that due to the structure of an annual programmatic review, ongoing improvement-oriented changes are made to the content but also taking into consideration the successive cohorts of learners.

I never recognized the decline in perceptions of gaining crisis resource management from the simulator sessions before I looked at the graphic. I’m not sure if it’s the course itself because that part of the course has not been modified all that much. I think what has changed though, is that everybody now comes to us competent in crisis management. I think that that’s what’s changed. Definitely. There’s not a resident who hasn’t heard about communication, call by name, close the loop, all the things we teach. Nobody was doing simulation before but now it’s introduced to students much earlier in their training. You’ve just proven to me that having a better evaluation process would be useful. If everybody hates simulation one year, well we’ll pick it up but if people’s ratings go down gradually… if everyone was giving it a five out of five, and two years later it is down to 4.2 it is still high but you need to look at the longitudinal trend to pick it up.

Moreover, this example provides evidence that the organization’s evaluation processes contributed to sustainable evaluation practice as ACES leaders have become even more motivated to collect additional program level data in order to keep pace with new trends (eg, earlier exposure to simulators in medical training) in medical education.

Resources dedicated to evaluation

More recently, ACES was acquired by the RCPSC. National ACES leadership explained that the new organizational structural stability was indeed beneficial in terms of resources. That is, leaders spoke about being less constrained from a resource perspective to delve into things like evaluation activities:
Now that we are with the Royal College [of Physicians and Surgeons], the beauty is that we have full time employees. So, if we decide that we’ll spend, let’s say, two weeks working on the evaluation, then we’ll just make sure that we organize schedules accordingly.

We will improve our performance as both an enabler and supporter of research and evaluation, so that our activities in this arena are better aligned with our mission. Among our activities will be to establish a more robust infrastructure for accessing research evidence, partner with leaders in research to leverage innovation, and establish more effective tools for assisting fellows in teaching and assessing the CanMEDS roles.

The more recent structure of the ACES organization within the RCPSGC has enabled leadership to plan to do more evaluative work on an annual basis with the reassurance that resources will be devoted to evaluation activities.

Discussion

The ECB literature provided grounding in the relevant strategies to examine ECB in practice within the national ACES organization. Specifically, the elements that guided our investigation included evaluation knowledge, skills and attitudes, and sustainable evaluation practice. Sustainable evaluation practice included subelements such as use of evaluation findings, shared evaluation beliefs and commitment, evaluation frameworks and processes, and resources dedicated to evaluation.1,2,6,8,9

Perhaps the strongest driving force for building ECB within the ACES organization comes directly from the internal motivation of leadership.20–22 That is, senior leadership was intrinsically motivated to engage in and sustain ECB efforts. This trigger for organizational ECB produced ongoing engagement between leadership and staff within the organization. A well-supported organization is best positioned for staff to have increased confidence to undertake evaluation activities.2 Moreover, the increase in engagement serves to increase the organization’s commitment to evaluation and their shared beliefs are central to improve the organization’s overall mission.23

Shared values within the organization, namely in involving others in interpreting and engaging with data, were found to create sustainable evaluation practice.7,9 That is, when groups of people in a system have intimate knowledge of data and have discussed its meaning and applicability, they have a possibility of developing a shared purpose and working together to reach organizational goals. Our study revealed that clinical instructors expressed positive emotions by being part of the ACES organizational network, and such engagement appears to be strongly linked a genuine desire to do more and learn more.

Many of the ECB strategies are interrelated and overlapping. The organization must have leadership who is intrinsically motivated to employ and use evaluation data to drive ongoing improvements within the organization.20,24 This includes risk taking when evaluation practices and processes may be somewhat foreign. That is, even though program evaluation may not be well understood by program leadership, they took a risk and plunged into working with the evaluative data. Overall, ACES leadership understands that evaluation is not something that is “added to the end” but rather incorporated in the organization’s operations.

With the appropriate mix of internal motivation, structure, and capacity, an organization can promote and maintain a culture of evaluative practices, specifically data use, for continuous improvement.25 Undeniably, creating and maintaining a culture of data use and ongoing inquiry have enabled national ACES to achieve a sustainable evaluation practice. While program leaders note that more work needs to be done, they have been able to put in place processes to promote ongoing evaluative activities.

Limitations

There are several limitations to our study. It was a cross-sectional study, undertaken within a single organization. The focus was placed at the leadership/instructor level, and while we feel that we were able to capture many important elements of ECB activity, we acknowledge the need to include program staff. Moreover, the national ACES organization may be unique in that leaders and instructors share close personal ties, as such findings here may not be easily transferrable to other organizations. Despite the potential unique context of this organization, we have been able to add to empirical literature that key elements contribute to organizational success using and learning from evaluative practices.

Conclusion

This study is important for medical organizations as they move toward increased accountability and in their practices and processes. We feel that we learned many valuable lessons by looking at ECB strategies in practice. Most notably, we found that organizational leadership must be motivated to use evaluation practices and processes in an ongoing manner. We are encouraged by the evaluation field’s research literature, models, and frameworks around ECB. Areas for
further exploration from this study include expanding the number of ECB strategies to include transfer of knowledge and exploring the extent of knowledge diffusion throughout the organization. It would be worthwhile to investigate the extent that ECB is different for leaders/instructors versus program staff (eg, administrative). Finally, on a larger scale, we are interested in what approaches and methods would be most meaningful and feasible for judging the effectiveness and impact of ECB activities.

Disclosure
The authors report no conflicts of interest in this work.

References