Measuring Evaluation Capacity in Ontario Public Health Units

Isabelle Bourgeois
École nationale d’administration publique

Nikolas Hotte
Eastern Ontario Health Unit

Louise Simmons
Eastern Ontario Health Unit

Raimi Osseni
École nationale d’administration publique

Abstract: This article presents a study of organizational capacity to do and use evaluation, conducted in 32 public health units in the province of Ontario. Methods include an organizational self-assessment using an instrument developed by Bourgeois, Toews, Whynot, and Lamarche (2013) as well as key informant interviews. Overall, our findings point to the fact that evaluation capacity is still developing in Ontario public health units; factors that support evaluation capacity in these organizations include the presence of an organization-wide evaluation policy, the availability of full-time evaluation staff in a supporting role, greater staff involvement in evaluation, and a standardized evaluation process. These findings highlight the importance of organizational structures and systems to evaluation utilization and provide potential areas of improvement for organizations wishing to improve their evaluation capacity.

Keywords: evaluation capacity, evaluation utilization, public health

Résumé : Cette étude vise à mesurer la capacité organisationnelle en évaluation de 32 bureaux de santé publique en Ontario. Pour ce faire, deux méthodes ont été employées : l’Instrument d’autoévaluation de la capacité organisationnelle en évaluation, élaboré par Bourgeois, Toews, Whynot et Lamarche (2013), ainsi que des entretiens semi-dirigés. Les résultats de l’étude démontrent que la majorité des bureaux de santé publique recensés développent toujours leur capacité à produire des évaluations de qualité et à les utiliser. Plusieurs facteurs organisationnels contribuent à renforcer cette capacité, tels que l’existence d’une politique organisationnelle d’évaluation, la disponibilité d’évaluateurs à temps plein pour diriger les...
travaux d'évaluation effectués par les programmes, l'implication directe du personnel dans la démarche évaluative, et une démarche uniformisée d'évaluation, appliquée de façon systématique. Les résultats de l'étude mettent en valeur l'importance des structures et systèmes organisationnels à l'utilisation de l'évaluation et suggèrent certaines pistes d'amélioration pour les organisations qui souhaitent renforcer leur capacité en évaluation.

**Mots clés** : capacité organisationnelle en évaluation, utilisation de l'évaluation, santé publique

In the province of Ontario, the delivery of public health programming is ensured by 36 public health units (PHU). Although each health unit is managed independently of the provincial Ministry of Health, all 36 units must respect common program standards and practices, including a requirement to “monitor program activities and outcomes to assess and improve the implementation and effectiveness of programs and services … [and] to facilitate public health practitioners' and policy-makers’ awareness of the factors that contribute to program effectiveness” (Ministry of Health and Long-term Care [MOHLTC], 2014, p. 26).

Although these evaluation requirements apply to all 36 Ontario public health units, important contextual and organizational differences exist between these organizations (e.g., rural vs. urban settings, population size and type, organizational size, etc.). These differences have, over the years, shaped how evaluation is conducted in each PHU in terms of responsibilities for evaluation as well as evaluation processes. They have also influenced each health unit’s capacity to conduct and use evaluation; therefore, a one-size-fits-all approach to evaluation capacity building (ECB) across all 36 organizations is not likely to be successful. Health units must therefore first have an understanding of their own current organizational evaluation capacity (EC) to select, design, and implement an appropriate ECB strategy. This article describes an empirical study conducted to select, adapt, and apply an organizational EC measurement instrument to Ontario health units to obtain a clear baseline of current evaluation capacity and to further support the identification of appropriate, customized ECB strategies for each participating PHU.

**BACKGROUND**

Evaluation capacity building generally refers to an organization’s capacity to do and use evaluation. Several authors have identified the ways in which organizations can improve the quality of evaluations as well as their use. For example, the development of organizational EC is known to be influenced by leadership, organizational environment, skill and knowledge development, policy development, resource allocation, and external supports (Carden & Earl, 2007; Carman & Fredericks, 2010; Cousins, Goh, Elliott, & Bourgeois, 2014; Garcia-Iriarte,
Several conceptual and empirically derived frameworks have been proposed by researchers to better organize and categorize our understanding of evaluation capacity (e.g., Bourgeois & Cousins, 2013; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; Preskill & Boyle, 2008). For the most part, these frameworks tend to reflect similar components of EC (see Bourgeois, 2016, for a comparison of five frameworks of EC). Some of them have in turn yielded measurement models and instruments that enable organizations to situate their current evaluation capacity against a series of set criteria or standards, and develop appropriate ECB strategies (see for example: Bourgeois, Toews, Whynot, & Lamarche, 2013; Nielsen, Lemire, & Skov, 2011; Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, & Balcazar, 2013).

RESEARCH OBJECTIVES

Two main objectives guided the present study: first, before undertaking any significant evaluation capacity building initiatives within Ontario public health units, it was necessary to measure each organization’s current level of evaluation capacity. Through such a baseline measure, we were hoping to enable each participating health unit to identify specific areas in which ECB initiatives are most needed (e.g., logic modelling) and, therefore, most likely to have a notable impact. Second, the aggregate measure provides an overview of evaluation capacity across all Ontario public health units and, therefore, may help identify general areas where more systematic interventions may be required in all 36 organizations.

CONCEPTUAL FRAMEWORK

Prior to undertaking this project, a team of evaluation professionals from several public health units conducted a literature review meant to identify key EC frameworks and measurement instruments. This review led the team to select an evaluation capacity framework proposed by Bourgeois and Cousins (2013), and this choice brought together the public health and academic research team members. This framework identifies six dimensions of EC; three of these dimensions illustrate an organization’s capacity to do evaluation (Human Resources, Organizational Resources, and Evaluation Planning & Activities) while the other three dimensions focus on an organization’s capacity to use evaluation (Evaluation Literacy, Organizational Decision-Making, and Learning Benefits). Each of these six dimensions is further divided into more specific subdimensions (19 subdimensions in total). Each subdimension is described qualitatively in a $19 \times 4$ matrix using four levels of capacity: low, developing, intermediate, and established. A bird's-eye view of the framework is depicted in Figure 1.
This conceptual framework was further operationalized into an organizational evaluation capacity measurement instrument by Bourgeois et al. (2013). This instrument was used to collect the data summarized below and is described in more detail in the next section.

**METHODOLOGY**

A descriptive, multicase, nonexperimental research design was used in the present study. Thirty-two of the 36 public health units participated in the study and were recruited through ongoing interactions with the study team. Ethical overview of the project was provided by the Eastern Ontario Health Unit using the Ethics Risk
Screen Tool provided by Public Health Ontario, as well as the ethics committee of the academic partner.

Two main research methods were used in this study: first, each participating health unit was asked to complete the Organizational Self-Assessment Evaluation Capacity Instrument (Bourgeois et al., 2013). Second, once the instrument was completed and submitted, the research team conducted a key informant interview with a representative from each participating health unit to collect contextual information to support the interpretation of the results obtained through the instrument.

**Instrumentation and Procedures**

The Organizational Self-Assessment Evaluation Capacity Instrument (hereinafter referred to as “the Instrument”) is based on the EC framework referenced above and measures, on two 4-point Likert scales (1 to 4), the current state of an organization’s evaluation capacity. The Instrument was developed in Microsoft Excel and features the same 19 subdimensions as those identified in the Bourgeois and Cousins (2013) EC framework. Each of these subdimensions is operationalized in the Instrument using more specific items, or statements, that provide the foundation for the self-assessment. The Instrument then automatically calculates a mean score for each subdimension, which is rolled up into a mean score for each dimension. The means are translated into the four capacity levels found in the EC framework (i.e., low, developing, intermediate, and exemplary or established) and their original descriptions.

The original EC framework and derived Instrument were based on an empirical study conducted with the Canadian federal government. Therefore, revisions were required to adapt some of the terms and concepts used in the Instrument to the provincial public health context. Two activities were conducted as part of this process: first, a thorough review of the Instrument was conducted by public health and academic team members, who provided written comments on potential areas of confusion or difficulty. These comments were discussed among team members, and modifications were made to the Instrument accordingly. Second, the Instrument was pilot tested in four health units—the results of the pilot test were captured through written feedback and key informant interviews and further modifications were made to the Instrument as a result of the pilot. These modifications largely consisted of changes in terminology and the removal of items that do not apply to the public health context. In addition to these preliminary adjustments, one of the two rating scales included in the original instrument (which measured the importance/priority/relevance of each item to the specific organizational context) was removed, as the terms were deemed to be confusing and unnecessary for public health units. This scale was replaced by a “not applicable” option, which enabled the removal of the item from the calculation of the mean score per subdimension. The final version of the Instrument contained 85 items in total.

As stated elsewhere (Bourgeois et al., 2013), the main purpose of the Instrument is to enable organizations to assess their own evaluation capacity to improve
it through the implementation of targeted strategies. The objective, therefore, was not to collect statistical information on the Instrument itself (for instance, reliability or internal consistency); this will be done through a separate study. Due to the self-assessment focus of our study, health units were encouraged to complete the Instrument in a small group setting that included key evaluation stakeholders in the organization (e.g., evaluation staff and managers, as well as program managers and senior managers). Although achieving consensus across all 85 items is typically a lengthy process (most consensus meetings took approximately three hours), it enables a broader view of organizational capacity and provides a structured forum in which certain issues related to the assessment could be discussed openly. According to several participants, these rich discussions added considerable value to the assessment process and contributed to increasing group members’ understanding of the challenges faced by evaluation professionals in the organization. In four cases, other approaches than the group consensus workshop were used to complete the Instrument for operational reasons. Each participating PHU was asked to complete the Instrument within 30 days. For each dimension and subdimension, a mean score between 1 and 2 out of a possible 4 can be interpreted as low capacity, a mean score between 2.1 and 2.6 can be interpreted as developing capacity, a mean score between 2.7 and 3.2 can be interpreted as intermediate capacity, and a mean score between 3.3 and 4 can be interpreted as established capacity. It should also be noted here that the completed Instrument for each PHU was sent to the study team directly; the results from each PHU are not publicly available and they were not shared between health units due to the study’s focus on organizational improvement. The overall means and medians obtained per dimension are presented in Table 1.

Following the submission of the completed Instrument, key contacts (evaluators, epidemiologists, etc.) in each PHU were asked to participate in a semistructured interview focusing on the broader organizational context for evaluation, evaluation policies and procedures in the organization, evaluation products and reports, and evaluation use. These interviews provided a complementary,

| Table 1. Overall Means and Medians per Dimension (score out of 4.00) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Dimension       | Capacity to Do  | Capacity to Use |
|                 | Human resources | Organizational resources | Evaluation planning and activities | Evaluation literacy | Organizational decision-making | Learning benefits |
| Group mean      | 2.75            | 2.57            | 2.59            | 2.91            | 2.55            | 2.67            |
| $n = 32$        |                 |                 |                 |                 |                 |                 |
| Group median    | 2.66            | 2.59            | 2.60            | 2.90            | 2.58            | 2.67            |
| $n = 32$        |                 |                 |                 |                 |                 |                 |
qualitative counterbalance to the results reported through the Instrument and enabled a richer analysis and description. The interviews took place in December 2014 and were conducted by telephone by two members of the research team. The interviews were not recorded, but detailed notes were taken by one team member while the other conducted the interview. The interview notes were then member-checked and analyzed by health unit and by theme; the data were used mainly to better interpret the findings obtained through the Instrument.

FINDINGS

Classification of Public Health Units
Each participating PHU was first classified into one of four levels of capacity (i.e., Low, Developing, Intermediate, or Established), based on a compound score for Capacity to Do (first three dimensions) and Capacity to Use (last three dimensions). The organizations were further divided into two groups for each level of capacity, based on whether they obtained a higher mean score for the “capacity to do” or “capacity to use” dimensions. Therefore, each PHU could be classified into one of eight possible groups: (a) low capacity—stronger capacity to do, (b) low capacity—stronger capacity to use, (c) developing capacity—stronger capacity to do, (d) developing capacity—stronger capacity to use, (e) established—stronger capacity to do, and (f) established—stronger capacity to use. The distribution of the health units, as well as the average scores for each subdimension, are shown in Table 2. None of the participating health units were classified as having “low capacity” across both sets of dimensions.

Once the capacity groupings were completed, we reviewed the instrument results for each PHU and conducted a content analysis of the interview findings by group. This analytical process yielded several themes, which are further described below.

Organizational Size
Key informants provided us with the approximate number of employees and/or full-time equivalencies (FTE) for their respective PHU. Overall, no link could be detected between organizational size and capacity-level groups. However, health units in the two Established groups were able to more clearly identify specific individuals or organizational roles related to evaluation as compared to some members of the Developing and Intermediate groups.

Organizational Structure
Our key informant interviews focused on how evaluation services are delivered in each organization. We discovered that evaluation can be delivered by a centralized unit of evaluation professionals, it can be assigned to program staff in a decentralized model, or it can be a blend of both, where program staff are primarily responsible for designing and implementing evaluations, with some support
Table 2. Average Dimension Score by Group (out of 4.00)

<table>
<thead>
<tr>
<th>Group</th>
<th>Capacity to Do</th>
<th>Capacity to Use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human resources</td>
<td>Organizational resources</td>
</tr>
<tr>
<td>a) Developing Stronger Do</td>
<td>2.64</td>
<td>2.26</td>
</tr>
<tr>
<td>(n = 9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Developing Stronger Use</td>
<td>2.39</td>
<td>2.22</td>
</tr>
<tr>
<td>(n = 7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Intermediate Stronger Do</td>
<td>3.07</td>
<td>3.13</td>
</tr>
<tr>
<td>(n = 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Intermediate Stronger Use</td>
<td>2.83</td>
<td>2.69</td>
</tr>
<tr>
<td>(n = 9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) Established Stronger Do</td>
<td>3.16</td>
<td>3.39</td>
</tr>
<tr>
<td>(n = 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f) Established Stronger Use</td>
<td>3.26</td>
<td>3.17</td>
</tr>
<tr>
<td>(n = 2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

from a core group of evaluation experts. This finding prompted us to consider the Involvement/Participation subdimension of the Instrument in relation to organizational structure. The group scores presented in Table 3 indicate that the groups with stronger Use scores tend to have a higher degree of staff involvement/participation in evaluation (with the exception of the minimal difference observed within the Established groups).

Organizational Frameworks and Policies Related to Evaluation

The “Organizational Evaluation Planning and Activities” subdimension includes assessment items related to organizational frameworks, policies, and other tools meant to structure evaluation activities within the organization. The mean scores for each grouping, provided in Table 4, indicate that there is a difference between the Developing and Intermediate groups for this subdimension but that the “Do” and “Use” distinctions do not seem to apply here.

The data captured during the organizational profile interviews corroborate the assessment instrument results. Most of the Developing health units do not have formal evaluation policies and procedures, whereas the majority of the
**Table 3.** Mean Scores for Involvement/Participation in Evaluation Subdimension (out of 4.00) and Key Interview Findings

<table>
<thead>
<tr>
<th>Analysis group</th>
<th>Involvement/participation in evaluation</th>
<th>Interview findings</th>
</tr>
</thead>
</table>
| a) Developing Stronger Do \((N = 9)\) | 2.83 | • Tendency toward centralized structures in many cases, although some hybrid models also observed: “Sometimes program staff conduct or assist with evaluations under the supervision of foundational standards staff.”  
• Involvement of program managers in evaluation is only mentioned a few times, and does not seem to be critical in conducting evaluations. |
| b) Developing Stronger Use \((N = 7)\) | 3.02 | • Tendency toward decentralized structures: “We don’t have anyone dedicated to evaluation.”  
• Some respondents in this group indicated that evaluation findings tend to be used immediately, which could be related to their propensity toward decentralization. |
| c) Intermediate Stronger Do \((N = 4)\) | 2.97 | • Responsibilities seem to be shared in this grouping, although in many cases evaluations are overseen by a central team or individual: “The research and evaluation committee is in charge of evaluation.”  
These may be hybrids with significant technical oversight. |
| d) Intermediate Stronger Use \((N = 9)\) | 3.23 | • Responsibilities also seem to be shared in this grouping, although there are variations in how evaluation is implemented.  
In one case, for example, one individual is responsible for evaluation and works with one other program staff member. In other cases, a few central epidemiologists/evaluators work with a small number of program staff. In still other cases, program staff mainly conduct evaluations with some oversight from central staff. |
| e) Established Stronger Do \((N = 1)\) | 3.75 | • In this particular health unit, a small team oversees evaluations and involves program staff when needed. |
| f) Established Stronger Use \((N = 2)\) | 3.73 | • These units deliver evaluation services through more centralized structures that also include program staff participation. |
Intermediate health units reported the existence of an organizational framework, policy, and/or procedures regarding evaluation. The Established PHUs also had varying forms of frameworks, policies, or procedures. Of note, the Established Stronger Do PHU has a specific research and evaluation policy that also includes the requirement to conduct at least one evaluation per program per year.

**Evaluation Process**

Along the same lines, we also explored whether each health unit followed a standardized evaluation process. Although the instrument did not specifically focus on this issue, the key informant interviews revealed interesting findings. In general, the two Developing groups were split as to whether or not they had a process for evaluation. In many cases, the respondents referred to the existence of a general evaluation process without clear guidelines. All PHUs in the Intermediate groups (both Do and Use) indicated that they had a process for evaluation. The Intermediate Stronger Do group indicated a more clearly defined process, however, than the Intermediate Stronger Use group, with clear responsibilities. Both Established groups were able to provide a description of a detailed process for evaluation. Based on these findings, it seems as though putting in place a detailed evaluation process that includes a description of key responsibilities reflects higher levels of evaluation capacity.

**Constraints and Challenges**

The subdimensions of Staffing, Leadership, and Budget are the most relevant when it comes to constraints related to organizational evaluation capacity. The average score per group for each of the previously mentioned subdimensions is provided in Table 5. All groups identified insufficient time and resources dedicated to evaluation. As the capacity level increases, however, it seems as though the specificity of the constraints also increases; for instance, higher capacity organizations will highlight a lack of skilled staff instead of simply a lack of human resources, or a need for local data instead of a need for more financial resources.

<table>
<thead>
<tr>
<th>Group</th>
<th>Organizational evaluation planning subdimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Developing Stronger Do (N = 9)</td>
<td>1.74</td>
</tr>
<tr>
<td>b) Developing Stronger Use (N = 7)</td>
<td>1.90</td>
</tr>
<tr>
<td>c) Intermediate Stronger Do (N = 4)</td>
<td>2.92</td>
</tr>
<tr>
<td>d) Intermediate Stronger Use (N = 9)</td>
<td>2.59</td>
</tr>
<tr>
<td>e) Established Stronger Do (N = 1)</td>
<td>3.33</td>
</tr>
<tr>
<td>f) Established Stronger Use (N = 2)</td>
<td>2.59</td>
</tr>
</tbody>
</table>
Dissemination and Uses of Evaluation

Many of the Developing health units did not identify clear dissemination mechanisms for evaluations (e.g., “no formal process to communicate findings” and program managers are “not asking for it, but they might be interested if they started to receive some of them”). In a few cases, some health units disseminate evaluation results, but only within their own department or sector. Generally, Intermediate and Established health units were able to identify intentional evaluation dissemination mechanisms and products. For example, these health units produce specialized reports, formal reports, “two-page reports for broader dissemination,” and infographics. They may also host knowledge exchange symposia, meetings, and presentations with external stakeholders. Of particular note, the Intermediate Stronger Use and Established groups—those with the highest overall evaluation capacity—reported most often that they tailored their communication products to a target audience (e.g., “the idea is to be more responsive to what we need and what our stakeholders need”).

The dissemination of evaluation findings is typically a first step in fostering evaluation use. The Instrumental/Conceptual Use and Process Use subdimensions measure the extent to which evaluation is shared and used by stakeholders. As shown in Table 6, the differences appear to be more distinct between capacity groups, rather than the Do/Use dichotomy, with the exception of process use for the Intermediate groups. Although the mean scores for Developing and Intermediate are within a fairly close range, there is an important jump in the case of Established PHUs.
The key informant interviews provided further details on how evaluation results are disseminated and used by health units in various groups.

**Developing**

Most Developing health units mentioned that evaluation findings are used to verify client satisfaction rather than monitoring program outcomes. In a few cases, staff used evaluation results to change program delivery and improve processes (i.e., change the “content and time of a course”). In addition, many of the health units in this capacity level grouping mentioned that they have limited resources to act on evaluation recommendations, and political pressures and public health expertise typically trump evaluation findings. Evaluation tends to serve accountability purposes and is not used in broader organizational decision-making: “I’m not sure if evaluation is really highly valued amongst staff … a box checked,” and “I don’t really have confidence that they read the report that I send to them … I don’t think that they really read or use them.”

**Intermediate**

Overall, Intermediate health units reported more positive findings related to the use of evaluation findings (e.g., discussion of comparability year over year, focusing on trends). Some respondents in the Intermediate-Stronger Do group mentioned that evaluation is not the only source of decision-making, but it is considered nonetheless. The Intermediate-Stronger Use group provided us with some clear examples of evaluation use: inform future program planning, improve existing programs, stop existing programs, pilot programming, and so on. The use of evaluations in budget discussions is mentioned several times by the Intermediate-Stronger Use group. Some health units within this group also mentioned that evidence is considered in planning evaluations. For example, “Every evaluation requested includes a question—what would you do with it?” This suggests that potential use is central in the planning stages of evaluations for this group.

<table>
<thead>
<tr>
<th>Analysis group</th>
<th>Instrumental/conceptual use</th>
<th>Process use</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Developing Stronger Do (N = 9)</td>
<td>2.51</td>
<td>2.23</td>
</tr>
<tr>
<td>b) Developing Stronger Use (N = 7)</td>
<td>2.56</td>
<td>2.45</td>
</tr>
<tr>
<td>c) Intermediate Stronger Do (N = 4)</td>
<td>2.75</td>
<td>2.39</td>
</tr>
<tr>
<td>d) Intermediate Stronger Use (N = 9)</td>
<td>2.95</td>
<td>2.71</td>
</tr>
<tr>
<td>e) Established Stronger Do (N = 1)</td>
<td>3.25</td>
<td>3.00</td>
</tr>
<tr>
<td>f) Established Stronger Use (N = 2)</td>
<td>3.38</td>
<td>3.22</td>
</tr>
</tbody>
</table>
Established  
The interview findings for this group were comparable to the Intermediate-Stronger Use group. One contact clearly expressed that “we do evaluations to inform decisions … they are taken seriously.”

DISCUSSION  
This study provides an overview of evaluation capacity in Ontario public health through the use of a common measurement tool. It moves beyond the more typical case narrative to compare and contrast various aspects of evaluation capacity between organizations with similar mandates and environments. Overall, we found that evaluation capacity is developing in most health units, although some organizations manifest higher levels of capacity to do and/or capacity to use evaluation. Our groupings into three levels of capacity (developing, intermediate, established), further broken down into two types of capacity (capacity to do and capacity to use), have enabled us to detect certain trends in organizational evaluation capacity. Organizations that show stronger capacity to do evaluation, for example, tend to have centralized evaluation structures staffed by technical experts, while organizations that show stronger capacity to use evaluation are generally more decentralized and involve program staff in the entire evaluation process. These findings corroborate others gathered through reflective narratives and organizational case studies (e.g., see Cousins & Bourgeois, 2014). Although this is not surprising, considering the theoretical contributions of the evaluation utilization and participatory evaluation literatures (see notably Cousins & Whitmore, 1998; Patton, 2008), our findings provide further context and options for organizations interested in increasing their evaluation capacity. Along the same lines, health units that have developed evaluation-related frameworks, policies, and procedures tend to achieve higher evaluation capacity levels than those whose evaluation work is less standardized. This may be due to the fact that such frameworks and tools provide a means through which higher quality can be achieved as well as a direct resource for new staff facing their first evaluation projects. Finally, the issue of use was explored through two different measures: first, the qualitative measure of evaluation dissemination, which showed that organizations who focus on producing different types of reports to meet various audiences’ needs tend to foster greater evaluation use; second, that the mean scores for the use subdimensions tend to reflect overall evaluation capacity groupings.

The findings collected through the Instrument were further developed based on our key informant interviews. As stated previously, the purpose of these interviews was to collect qualitative descriptions of each health unit to better contextualize and understand the Instrument results. Taken together, the quantitative and qualitative findings have enabled us to reach certain conclusions regarding organizational evaluation capacity and link these back to the EC literature.
Organizational Size

In our study, organizational size did not appear to influence evaluation capacity level. As noted previously, health units of all sizes were found among the Developing, Intermediate, and Established groups. This was somewhat surprising—we had expected some influence of size on EC, based on previous studies such as the Kegeles, Rebchook, and Tebbetts (2005) paper that commented that smaller community-based organizations, “which have fewer financial and staffing resources, seemed to perceive evaluation to be a burden that detracted from programming” (p. 293). Our findings instead seem to support Baron (2011): “no matter what the size of the organization, internal evaluation can be a prevalent, thriving activity that reaches beyond the political, economic, or social barriers to propel the organization forward” (p. 88).

Organizational Structure

The location of the evaluation function within the organization has long been recognized as playing an important role in support of EC (see, for example, Bourgeois, Hart, Townsend, & Gagné, 2011). Centralized evaluation structures tend to have a few staff members who carry most, if not all, of the responsibility for evaluation (Bourgeois et al., 2011). Decentralized structures tend to have all of the evaluation activities conducted at the program level. Our findings, however, show that very few health units can be squarely situated in either one of these structures. In most cases, we found hybrid structures that combined some of the centralized technical support, coupled with decentralized program staff evaluation responsibilities. As mentioned previously, Developing and Intermediate health units that scored higher for “capacity to use” tend to have greater staff involvement in evaluations. Program staff participation in evaluations allows for a better understanding of the program context; thus, it is likely that findings provided by such an evaluation are more applicable (Patton, 2008). In addition, key stakeholders that participate in an evaluation will most likely have a better understanding of the findings and recommendations; therefore, it is possible that they will buy into and apply the recommendations, sometimes without waiting for the evaluation to be completed (Brazil, 1999; Dabelstein, 2003; Patton, 2008). This is also in line with the opposite finding, that those health units that scored higher for “capacity to do” tend to have more centralized structures. The in-depth knowledge, competencies, and experience of centralized evaluation practitioners enable the organization to produce quality evaluations with rigour and credibility. An interesting finding relates to the three Established health units: in all these cases, a centralized structure ensured oversight over the evaluation function, while still maintaining a high level of stakeholder involvement. These health units may be good examples of successful hybrid structures. Regardless of these results, however, the notion of centralized/decentralized/hybrid structures and their influence on EC requires further study, to identify the specific conditions that allow us to successfully draw on the strengths of both structures.


**Frameworks, Policies and Evaluation Process**

Formal evaluation frameworks, policies, and processes refer to the organizational requirements and supports for evaluation. For instance, such documents can include evaluation budgets, roles and responsibilities regarding evaluation, process for proposing/approving evaluations, ethics review process, communication of evaluation results, and data monitoring systems.

As previously mentioned, the distinction between the Developing and Intermediate/Established groups is quite noticeable with regards to this component. Clearly, organizations that have given more extensive thought to the integration of evaluation within their ongoing processes and procedures, and have formalized this integration through frameworks and/or policies, appear to have attained a greater degree of capacity overall.

**Constraints and Challenges**

As previously mentioned, a common theme among all groups is the lack of time and resources dedicated to evaluation, which is also highlighted throughout the literature (Cousins, Bourgeois, & Associates, 2014; Lennie, 2005; Ohmer, 2008). As knowledge of evaluation and its potential in supporting decision-making grows in an organization, so does demand. King (2002) reminds us to maintain balance “between creating demand and meeting expectations in a substantive and timely manner” (p. 77), which will limit possible frustrations and stimulate EC growth.

It is also interesting to note that the constraints identified seem to become more sophisticated as higher evaluation capacity levels are attained. Staff capacity, evaluation budget, and leadership are key components that can benefit or limit evaluation capacity (Cousins, Bourgeois, & Associates, 2014; Milstein, Chapel, Wetterhall, & Cotton, 2002). The grouped scores found within the Staffing, Budget, and Leadership subdimensions indicate that conditions may be more favourable in the Intermediate-Stronger Do and Established health units, which could explain why they have moved on to a higher specificity of need as stated in the results section.

**Communication of Findings and Uses of Evaluation**

Communication of findings to the appropriate internal and external stakeholders is an important aspect of evaluation that affects the utilization of evaluation (Patton, 2008; Rossi, Lipsey, & Freeman, 2004). Our findings point to a strong link between communication of evaluation findings and their use. The Intermediate-Stronger Use and Established health units that tailor communications products to a target audience appear to be fostering use in a more systematic and strategic manner than the other groups.

**LIMITATIONS OF THE STUDY**

One of the potential limitations of this study is the nature of the instrument itself. By choosing to build a self-assessment instrument, Bourgeois et al. (2013) wanted to focus on supporting organizations interested in measuring and developing
their own evaluation capacity, rather than comparing across different organizations. Therefore, the instrument items are likely to have been interpreted in different ways across health units; this means that any comparative analysis of the instrument results must be interpreted with caution; we consciously chose to not use comparative measures or tests of significance because of this. However, the general trends revealed through the grouped data analysis provide interesting information about capacity levels across all health units and support the development of ECB strategies for all organizations.

We also recognize that the group consensus process used to complete the instrument may temper some of the rankings toward the middle ratings instead of extremes. The four-point Likert scale was a conscious decision for the following reasons: groups may have a tendency to default to the middle answer, ratings correspond to the four possible levels of capacity found in the EC framework, and an abundance of choices may increase the amount of time to select a specific rating item within a group format. Groups were encouraged to access the EC framework if they had any concerns as to whether or not a rating was appropriate.

**CONCLUSION**

Overall, the current state of evaluation capacity among Ontario public health units can be classified as “developing” according to the Bourgeois et al. (2013) Organizational Evaluation Capacity Self-Assessment Instrument. The propensity for health units to be classified in this category is not surprising, given that evaluation-related requirements are relatively new and evaluation capacity may not have been a priority in many organizations before their implementation. Organizational evaluation frameworks and policies and a clear process for evaluations should contribute to improving EC, as should intentional dissemination and use strategies. The question of centralized/decentralized/hybrid structures within a PHU requires further study; however, an approach that maintains the rigour of evaluations and contributes to the use of findings should improve EC regardless of form or shape. These findings highlight the importance of organizational structures and systems to evaluation utilization and provide potential areas of improvement for organizations wishing to improve their evaluation capacity.

Based on the findings described here, our next steps are to develop a series of ECB strategies and to test them empirically in interested health units. Our baseline measure will be used to compare evaluation capacity levels in each health unit before and after implementation of specific ECB activities and organizational changes.

**ACKNOWLEDGEMENTS**

The Building Evaluation Capacity in Ontario’s Public Health Units LDCP team would like to thank Public Health Ontario (PHO) for its support of this project. The team gratefully acknowledges funding received from PHO through the Locally Driven Collaborative Projects program.
DISCLAIMER
The views expressed in this publication are the views of the project team, and do not necessarily reflect those of Public Health Ontario.

NOTE
1 The Instrument is not meant to be used as a survey, where several people within the same organization complete and submit their own copy of the Instrument: only one Instrument should be completed per organization. Because the Instrument was adapted to this particular set of organizations, only descriptive statistics are reported in this article; other data concerning reliability, internal consistency, and the covariance structure were not calculated as part of this specific project.

REFERENCES


**AUTHOR INFORMATION**

**Isabelle Bourgeois**, Ph.D., is an assistant professor at l’École nationale d’administration publique (ENAP). Her research work focuses on measuring organizational evaluation capacity and evaluation utilization. She is the scientific director of the Centre de recherche et d’expertise en évaluation (CREXE). She is also currently the Vice-Chair of the Consortium of Universities for Evaluation Education as well as the Associate Editor of the *Canadian Journal of Program Evaluation*.

**Nikolas Hotte**, BSc., MHS(C), is a health promotion specialist with the Eastern Ontario Health Unit. He was the coordinator for the Locally Driven Collaborative Project-Building Evaluation Capacity in Ontario Public Health Units.

**Louise Simmons** holds an M.Sc. in Sciences de l'activité physique from the Université de Montréal. She has worked at the Eastern Ontario Health Unit for the past 20 years, 17 as a program and office manager. Presently she is the manager of Planning and Evaluation, and the lead for the LDCP Program Evaluation project.

**Raïmi Osseni**, MPA, is an analyst at Treasury Board Secretariat of Canada’s Centre of Excellence in Evaluation (Program Performance and Evaluation Division) where he supports the implementation of the new Policy on Results across the public service. Raïmi also co-founded, led, and contributed to Évalue-action and to the Réseau Francophone des Évaluateurs Émergents (RF-Ée). He currently sits as treasurer on the Board of Directors of the Canadian Evaluation Society’s National Capital Chapter.