Government and Voluntary Sector Differences in Organizational Capacity to Do and Use Evaluation

J. Bradley Cousins
Faculty of Education
University of Ottawa

Swee C. Goh
TELFER School of Management
University of Ottawa

Catherine Elliott
Faculty of Education
University of Ottawa

&

Tim Aubry
School of Psychology
University of Ottawa

Presented at the annual meeting of the Canadian Evaluation Society, Québec, May 2008

Correspondence: B. Cousins, Faculty of Education University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, ON, CANADA, K1N 6N5, bcousins@uottawa.ca. The authors gratefully acknowledge financial support for this research provided by the Social Science and Humanities Research Council of Canada (#410-2005-0303) and technical assistance for web-based data collection provided by Réseau Circum Inc.
Increasingly evaluation scholars and practitioners have expressed a significant interest in understanding organizational capacity for evaluation (Barnette & Sanders, 2003; Cousins, Goh, Clark, & Lee, 2004). Conceptually, this interest lies at the confluence of three distinct, yet related, streams of inquiry. First, several authors have focused on evaluation capacity building: how to develop within organizations the knowledge and skills required to conduct or commission and receive evaluations that are methodologically defensible and relevant to organizational information needs. This stream of inquiry has a significant focus on training and development, including understanding beyond that which is represented by program evaluator competencies. (Boyle & Preskill, 2007; Preskill & Russ-Eft, 2005). A second domain of interest is evaluation utilization, which has captured the interests of evaluation scholars for over forty years (Cousins & Leithwood, 1986; Patton et al., 1977). In recent years investigators have turned their attention toward such questions as ‘how can evaluation use and influence be differentiated?’ (Kirkhart, 2000; Henry & Mark, 2004) and ‘what is the relative impact of use of evaluation findings as opposed to use of evaluation processes?’ (Cousins, 2007; Patton, 1997). Finally, authors have connected the consequences of evaluation to organizational development most notably organizational learning (e.g., Cousins & Earl, 1995; Forss, Cracknell & Samset, 1994; Owen & Lambert, 1995; Preskill, 1994).

Perhaps the most significant benefit of locating organizational evaluation capacity in these streams of inquiry is that conceptions of ECB have broadened from a focus on training and development to considerations about the organizational capacity to use evaluation (Cousins et al., 2004; Stockdill, Baizerman & Compton, 2002). Conceptually, integrating evaluation into organizational routines and culture is as much about building the capacity to do evaluation as it is to building the capacity to use it.

While conceptual advances have been made, research on organizational evaluation capacity is wanting. Much of what is known rests on findings emerging from post hoc narrative reports on evaluation or evaluation capacity building experiences as opposed to planned empirical inquiry using quantitative, qualitative or mixed-methods (Cousins et al., 2004). In this paper we report findings emerging from a broad-based program of research on evaluation with an explicit focus on organizational evaluation capacity. Specifically, we build on a pan-Canadian survey of internal evaluators (Cousins, Goh & Elliott, 2007) to examine differences in self-reported understandings of evaluation capacity of members of government and voluntary sector organizations. We were interested in discovering if and how perspectives between these groups differ with respect to the antecedent conditions underlying evaluation practice in organizations, the organizational capacity to do evaluation and the capacity to use it.

We define evaluation as systematic inquiry used in the formulation of judgments about an evaluand’s (e.g., program, strategy, innovation) merit, value or significance, or in support of decision making concerning the evaluand. It has long been understood that evaluation takes place within a political context (e.g., Chemlimsky, 1995; Mohan & Sullivan, 2006; Weiss, 1973) and that the extent to which evaluation is embraced by organizations will depend in part on organizational information needs and exigencies. Evaluation can serve a hard-nosed, judgmental, accountability function or it can support
learning through inquiry about evaluand strengths and weaknesses in an improvement-oriented manner.

Both government and voluntary sectors in Canada and elsewhere have experienced increasing demand for evaluation in recent years. Evaluation and performance measurement are natural elements of new public management systems with a focus on outcomes for society and results-based management. An increasing focus on outcomes in terms of management and funding is characteristic of funding agencies in the voluntary sector as well as public service management. Moreover, the recent legislation of the Federal Accountability Act in Canada has stimulated interest in evaluation in the government sector (Government of Canada, 2007).

While government and voluntary sector organizations share these information needs, context considerations are quite distinct, the former being large, heavily bureaucratic and openly subject to changes in government priorities through the political process, and the latter being typically small, front-line and subject to the demands of funding agencies (including government, foundations and other non-governmental funding agencies and networks). Understanding the context within which evaluation takes place is essential to understanding how evaluation is integrated into organizational culture. For this reason we are interested in examining the differences between government and voluntary organizations in terms of their capacity to do and use evaluation.

**EVALUATION IN GOVERNMENT VERSUS NOT-FOR-PROFIT ORGANIZATIONS**

In this section we briefly provide some background on evaluation in government versus that taking place in the voluntary sector. First, we consider the nature of the evaluation function within public management and how the function is managed by government.

**Government and Evaluation**

Davies, Newcomer and Haluk (2006) make the case that government has two primary interests in evaluation, namely the promotion of accountability (informing the public, decision makers, taxpayers, service users and other stakeholders about the worth of government policies, programs and interventions) and improving government management as part of good public management. Evaluation, results-based management, performance-based management, and the like are all seen as tools of new public management. In Canada, principles of new public management have been embraced by federal and provincial governments alike (Segsworth, 2005).

Government demand for evaluation arises principally from legislative and executive mandates (Davies et al., 2006) although evaluation through commission systems is also a longstanding focus. Through the establishment of procedures and policies connected with budgetary processes, government executives typically require evaluation as a means of exerting control over agencies. In Canada, similar to the case in the US as portrayed by Datta (2003), government has played a key role in creating the demand for evaluation by requiring that money be set aside for the evaluation of grants.
and contributions programs and more recently, by developing a federal management accountability framework, within which evaluation is situated.

Governments use a range of evaluation methods within the context of performance measurement (Davies et al., 2006; Gussman, 2005). According to Datta (2003), not only does government use a range of evaluation methods, but indeed it has served to shape the methods that are used. Impact evaluation which involves target setting and monitoring is essentially goals-based evaluation. One dimension of impact evaluation would be value-for-money performance management which uses cost-effectiveness analysis and other econometric tools. Impact evaluation relies on a variety of sources of evidence including experimental and quasi-experimental evaluation, systematic reviews and integration or ‘meta-analysis’ of existing evidence, statistical and econometric modeling. Another major thrust of evaluation in government is formative or process evaluation where the intent is to inform program and policy development and improvement on the basis of evidence.

While independent institutions such as universities and other private and public agencies have contributed to evaluation in government (Davies et al., 2006), the function is typically managed by auditing agencies, internal evaluation units of government and contract evaluators hired by governments. Globally, governments are increasingly establishing specific units of evaluation “thus bringing evaluation into the very cradle of government structure” (Davies et al., 2006, p. 173). In Canada by the mid 1990s program evaluation and audit were becoming less distinct as administrative functions with over 50% of federal departments placing evaluators and auditors in a combined unit, as compared with about one third in 1993 (McDavid & Huse, 2006).

Recent studies have brought into question the quality and relevance of evaluation in government in Canada, and ultimately its use (Breem et al., 2005; Gusman, 2005). In a recent study commissioned by the Centre of Excellence for Evaluation (CEE) at the Treasury Board of Canada, Secretariat (TBS) evaluation was found to be disconnected with senior management decision-making. Breem and associates (2005) interviewed Deputy Ministers and found that although they consistently expressed the view that evaluation is a policy/program function, they identified a lack of a feedback loop between evaluation findings and policy/program development and management. Deputies were also aware of and concerned about resource limitations, recognizing that the practice of contracting out evaluation potentially weakens the ability of evaluation units to become ongoing sources of advice and wisdom. Aucoin (2005) maintains that the enhancement of quality in public sector program evaluation will not happen unless there is the demand for quality from senior officials or ministers.

This situation is similar to that occurring in other jurisdictions including the US. VanLandingham (2006) and Perry, Thomas, DuBois and McCowan (2006) both lament the continuing tensions between the demand for independence, on the one hand, and the need to employ utilization-oriented strategies such as stakeholder engagement and relationship building, on the other. Despite Perry et al.’s (2006) report on some success of a performance auditing office in grappling with this problem, there is little doubt that the tension persists and that indeed it may be exacerbated by coupling audit and evaluation functions.

The onset of new public management and a focus on results-based management has fuelled once again the performance measurement industry in government. Despite an
obvious, potentially symbiotic role for evaluation in performance measurement, to a significant degree this domain of practice remains contested turf between evaluators and auditors (Mayne, 2006). Mohan, Tikoo, Capela and Berstein (2006) argue that evaluators have the expertise and tools to assist in the development of sustainable performance measurement systems and in so doing would ultimately benefit from the existence of such systems in terms of developing deeper understanding of performance criteria, expectations, and so forth. Mayne and Rist (2006) concur: they suggest that the evaluator’s role should be expanded to assisting in interpreting performance measurement information, using the information to strengthen the knowledge and learning function, and developing streams of evaluative knowledge, as opposed to single studies. Yet what is recommended and what is happening appear to be two different things at this juncture.

In summary, government has been a longstanding driver of evaluation in most jurisdictions, yet tensions between accountability and learning needs, and between stakeholder engagement strategies and demands for independence have hampered the usefulness of evaluation. We now lay out the landscape for evaluation in the voluntary sector by way of providing a comparative context for our study.

**Evaluation in the Voluntary Sector**

Similar to the functions of program evaluation in government, there have been increasing expectations on organizations in the voluntary by their funders to conduct program evaluation for the purpose of accountability and quality assurance (Bozzo, 2002; Chaytor, MacDonald, & Melvin, 2002; Juillet, Andrew, Aubry, & Mrenica, 2001). The changing funding environment in Canada in the voluntary sector -- where funding is allocated for discrete projects combined with shrinking resources -- has fuelled this increase in program evaluation activities in the sector (Bozzo, 2002; Hall, Phillips, Meillat, and Pickering, 2003; Juillet et al., 2001).

In reviewing the history of program evaluation in the voluntary sector in the United States, Plantz, Greenway, and Hendricks (1997) noted how the requirements evolved from initially auditing program expenses, (i.e., documenting how money was spent), in the 1970s, to documenting program recipients and program outputs or products in the early 1980s, to examining client satisfaction in the late 1980s, and finally to evaluating program outcomes in the 1990s. A major training initiative by the United Way of America intended to equip grantees with the skills required to measure program outcomes has served as a catalyst to moving the voluntary sector towards evaluating outcomes in the United States (Hatry, van Houten, Plantz, & Greenway, 1996). A similar evolution in this direction, (i.e., evaluation of program outcomes), is also now prevalent in the voluntary sector in Canada (Hall et al., 2003).

In order to document the program evaluation practices of and challenges faced by organizations in the voluntary sector in Canada, Hall, and his colleague (2003) conducted a national survey of 1,965 voluntary organizations in the areas of arts and culture, education, health, and social services and 322 funders from government and public and private foundations. Almost three-quarters of surveyed organizations indicated that they conducted program evaluations routinely. In line with the move to evaluating outcomes, 89% of the voluntary organizations who had conducted evaluations in the past year indicated that they were expected by funders to evaluate the outcomes or impacts of their
programs. Two-thirds of these organizations described themselves as having successfully evaluated program outcomes. However, the researchers found that there was confusion in the sector about the differentiation between outputs and outcomes with many of these organizations actually collecting data on outputs only (Hall et al., 2003).

In the context of increasing expectations to conduct program evaluation, organizations in the voluntary sector face a number of challenges. These include gaining support for program evaluation from staff and volunteers, integrating program evaluation into service delivery programs, capturing and measuring outcomes of voluntary programs, and having limited capacity in many organizations to conduct rigorous and useful evaluations (Bozzo, 2002; Hall et al., 2001; Cutt & Murray, 2000; Juillet et al., 2001; United Way of America, 2000).

Of these challenges, the lack of evaluation capacity of organizations in the voluntary sector has proven itself to be particularly limiting in terms of them being able to both do and use program evaluation as an organizational learning tool (Bozzo, 2002; Hall et al., 2001; United Way of America, 2000). This lack of capacity has involved a paucity of both internal skills and resources (Bozzo, 2002). For instance, Hall and his colleagues found that over half of voluntary organizations used their internal funds for program evaluation and that only approximately one in five (19%) of funders provided sufficient supplementary resources for organizations to conduct evaluations. Therefore, it was not surprising that only a very small number (6%) of organizations relied on external evaluators. There is consensus in the literature that the technical knowledge and skill level among personnel in the voluntary sector related to conducting program evaluations is low (Bozzo, 2002; Chaytor et al, 2002; Hall et al., 2001; Juillet et al., 2001).

As a result of this lack of capacity in the voluntary sector, there have been a number of efforts taken to increase capacity through a combination of didactic materials, training, and technical support at the program level (Compton, Glover-Kudon, Smith, & Avery, 2002; Stevenson, Florin, Mills, & Andrade, 2002) and at the system level (Atkinson, Wilson, & Avula, 2005; Naccarella et al., 2007; United Way of America, 2000). Research in this area suggests that internal capacity building is being accomplished at least in some areas of the voluntary sector. However, efforts in this area remain a work in progress and even organizations with personnel trained in program evaluation are faced with limited financial resources and available time for engaging in evaluation activities (Cutt & Murray, 2000; Hall et al., 2001; Juillet et al., 2001).

We now turn to an explication of the guiding conceptual framework and research objectives for the present study, followed by the methods and results.

**CONCEPTUAL FRAMEWORK**

Figure 1 depicts the conceptual framework that we developed to guide the research. The framework is very much grounded in our prior work (Cousins et al., 2004; 2005) including a recent concept mapping study inquiring into evaluation use in government (Cousins et al., 2006). We now describe the various constructs, all of which influenced the design of the questionnaire survey that we used.

---

1 The section adapted from Cousins et al. (2007)
Capacity for Organizational Learning

We think of evaluation capacity as being integral to the larger cultural construct of organizational learning capacity (OLC). To represent OLC we adopted Goh’s (1998) conceptualization of a learning organization. Those organizations with a higher organizational learning capacity are more likely to demonstrate five learning organization characteristics or building blocks, as defined in this model: clarity and support for mission and vision, leadership that supports learning, an experimenting organizational culture, the ability to transfer knowledge effectively, and teamwork and cooperation. These attributes are mutually supportive conditions, which foster a learning organization. The strength of each of these attributes is measured using the Organizational Learning Survey developed by Goh and Richards (1997). To the extent that evaluators see their organizations as having these attributes, then the greater the degree the organization will be seen as having a capacity for learning.

Organizational Support Structures

OLC depends in part on organizational support structures. Goh (1998) contends that the five organizational building blocks mentioned above -- mission and vision, leadership, experimentation, transfer of knowledge, teamwork and cooperation -- are dependent upon two supporting foundations. The first, organization design, refers to the fact that organizational learning is enhanced when an organization has a flat, non-hierarchical structure with few formalized controls over employees’ work. The second, employee skills and competencies, refers to the need for organizations to invest in professional development by building team competencies through shared learning experiences. Rather than focusing only on individual job-based skills, learning organizations emphasize the development of general behaviors such as critical thinking and collective problem solving skills (Goh, 1998). Other organizational support structures, which enhance OLC, include communication systems to enhance knowledge dissemination and appropriate recognition programs to reward and motivate employees. Since this may be critical to enhancing learning activities, we asked evaluators the degree to which they perceive these support structures to be present in their organizations.

Capacity to Do Evaluation

The capacity to do evaluation refers to all of the knowledge and skills required to carry out or oversee evaluation activities. These would include knowledge and skills about evaluation planning, ethical practice, instrument development, data collection, analysis and interpretation, and reporting and follow up.

Sources of Knowledge, Skill and Ability

Evaluation capacity building often depends on training and professional development opportunities. These would include on-the-job informal learning
experiences as well as more formalized opportunities such as workshops, courses, degree or certificate programs.

**Evaluative Inquiry**

Evaluative inquiry refers to a range of activities, from needs assessment, program evaluation and outcome monitoring to environmental scanning. Evaluative activities vary in their intent, their objects, the type of information collected and methodology employed. They may be “summative” in nature, providing judgment of the merit or worth of a program, or “formative”, providing information for decision making. Typically, a summative evaluation is performed upon completion of a program, for the purposes of accountability, certification, or selection (Nevo, 1983). In contrast, a formative evaluation is aimed at program improvement, and is typically conducted while the activity is ongoing or the program is developed. It may be performed by external evaluators, internal personnel, or a combination of both. A recent trend has been to increase the participation of stakeholders in evaluation activities, with the primary intent of enhancing the utilization of results; but within this “collaborative” approach, there is some disagreement among authors on the issues of participant selection, depth/type of involvement, and control of the evaluation technical decision making (Cousins, Donohue & Bloom, 1996; Turnbull, 1999).

Our own conceptualization of evaluation emphasizes the positive, formative aspects of evaluation, as described earlier. We believe that evaluative inquiry can assist decision makers in making informed decisions about an object such as an improvement program or project (Cousins & Earl, 1995). Questions related to this construct assess the degree to which evaluators see evidence of a wide range of evaluation activities in their organization.

**Capacity to Use Evaluation**

Apart from the supply or production side of organizational evaluative inquiry, we need to think about the capacity for demand-side organization members (users) to actually process and use evaluation findings or to benefit from evaluation in some other way. Use of findings would include support for organizational or program decision making, learning about the organization or program, or use of the evaluation in symbolic (e.g., compliance indicator) or political (e.g., persuasive) uses of results.

The capacity to use evaluation is also reflected in process use, or the extent to which demand side members of the organization or other non-evaluators benefit from being close to or participating in the evaluation. Process use might for example, imply the development of understanding of evaluation logic or research skills, an enhanced tendency to question basic assumptions or to benefit individually or collectively in terms of developing an evaluative habit of mind (Katz, Sutherland & Earl, 2002).

**Mediating Conditions**

Our research on what makes evaluation useful in government yielded some interesting ideas about factors or conditions that mediate the use of evaluation. Some of
these were evaluation quality, evaluator credibility, involvement of users, and perceived timeliness. Several of these relate to factors that have been reported in the literature for quite some time (see, e.g., Cousins & Leithwood, 1986).

**PURPOSE OF THE STUDY**

This study builds on the results from the survey of Canadian internal evaluators mentioned above (see Cousins et al., 2007). Our primary objective is to understand differences in self-reported perceptions and opinions about organizational capacity for evaluation from government versus voluntary sector perspectives. First we wanted to determine the extent to which internal evaluators for these two groups were similar or different with regard to background and demographic characteristics. Then we planned an examination of the differences between the perceptions of personnel involved in program evaluation in these sectors with respect to:

1. Contextual circumstances for organizational evaluation capacity;
2. Organizational capacity to do evaluation; and
3. Organizational capacity to use evaluation.

A deeper understanding of organizational evaluation capacity from the perspectives of government and voluntary sector internal evaluators will have significant implications not only for ECB theory and practice but also for evaluation policy development. This research is also likely to raise questions for future inquiry.

**METHODS**

The methods used to implement the Canada-wide survey are detailed by Cousins et al. (2007). In this section we provide a summary of the methods used to gather the data; the reader is referred to Cousins et al. (2007) for more detail.

**Procedure and Sample**

We were interested in recruiting internal evaluators or organization members who have responsibility for oversight of evaluation work that is contracted out. Through various evaluation and organizational networks – specifically CES, TBS, United Way of Greater Toronto and United Way of Ottawa – we invited members to participate in an online questionnaire survey about organizational evaluation capacity. Follow-up email reminders were sent. The survey was conducted in English or French depending on a participant’s preference.

In the end we achieved a sample of 340 respondents. It is not possible to estimate the response rate for our survey since we were unable to confirm how many internal evaluators actually received the invitation to participate (nor do we have a clear estimate of the Canadian population of internal evaluators). Approximately half of the respondents are from the public government sectors including federal, provincial and municipal, another 25% are from the non-profit private sector and the remaining participants are from a number of other types of organizations (e.g., private firms, colleges/universities). Most of the respondents worked as evaluators or performance measurement specialists (40%), and about 33% were either program managers or held senior administrative
positions. The remaining respondents (27%) either did not respond or specified another job category. Average experience working as in evaluation was 7.1 years and with about 50% of time estimated as being spent on evaluation activities. The observed level of knowledge of evaluation theory and practice was quite high and respondents were well educated with 95% having either an undergraduate or higher level graduate degree. Close to two thirds of the participants were female and about the same proportion were in their 40s.

For this study we restricted our analysis to persons belonging to government (N=160) and voluntary sector (N=89) for a total N of 249. Government respondents were mostly from the federal sector (59%) with 35% provincial and 5% municipal. Voluntary sector respondents were mostly from not-for-profit organizations (87%) with some from non-governmental organizations (13%).

Instrument

The questionnaire to capture all the constructs of interest in our study was developed as a hybrid instrument. There were eleven parts to the survey, the first and last asked about background information from respondents. Sections in-between measured self reported opinions about:

Contextual Circumstances
- organizational learning capacity;
- organizational support structures, including degree of formalization and training support and development to increase skills and knowledge of employees in an organization;
- the types of evaluation capacity building activities such as professional development or on the job training used to develop capacity, and

Capacity to Do Evaluation
- the organizational capacity to carry out or oversee the evaluation function as well as resource availability to support evaluation activities;
- the nature and frequency of evaluative inquiry within the organization; and
- the degree to which non-evaluator stakeholders are involved in doing evaluation.

Capacity to Use Evaluation
- use of evaluation findings and process and their consequences for the organization and individuals within;
- and conditions mediating evaluation use

Operationalization of Variables

Table 1 describes how we computed scale variables for each of the constructs of interest. For the most part, scale variables were linear combinations of subsets of questionnaire items pertaining to the various constructs in our conceptual framework. Respondents rated each item on a seven-point Likert type scale for either opinion (endpoints ranging from 1 ‘strongly disagree’ to 7 ‘strongly agree’) or frequency of behavior or practice (endpoints ranging from 1 ‘never’ to 7 ‘always’). In the table we
reproduce the top five highest rated items included in the linear combination. (For a complete listing of items and their distribution, see Cousins et al., 2007). We calculated Cronbach’s alpha as a measure of internal consistency for each of the scale variables.

--- Insert Table 1 about here ---

The first four variables in Table 1 are associated with the antecedent condition underlying the extent to which evaluation is integrated into the organizational management function. Note that we chose to include ‘organizational learning capacity’ (OLC) a cultural variable as part of this group although it is represented by a more direct link to the evaluation function in our conceptual framework. The next four variables are associated with the capacity to do evaluation, including a direct measure of capacity and associated measures of organizational evaluation activities, and two measures of stakeholder involvement in evaluation. Finally, we listed variables associated with the capacity to use evaluation. Included here are use of findings, use of evaluation processes and conditions mediating use within the organization.

In the Table we can see that virtually all of the variables were shown to be reliable with several showing alpha coefficients of .90 or higher. One variable – ‘organizational support structures (OSS), formalization’ – showed an alpha coefficient of only .68, which is comparatively low but acceptable for our purposes. We can also observe that average scores on the measures tended to be slightly higher than the scale midpoint of 4.5 in many cases, although associated with some variables were means that were either at or slightly less than the scale midpoint (e.g., OLC, OSS-formalization, stakeholder participation-level and frequency). Variability in responses across the scale variables was quite uniform with observed standard deviations of slightly more than 1.0 on average on the seven point scale. This implies that about two thirds of the responses were within the interval defined by 4.2 and 5.2 on the seven point scale. Scale variables were reasonable approximations to normal distributions given the range of observed N = 270 to 320.

**Plan for Analysis**

Our first step in the analysis was to better understand our sample in terms of organizational role by examining for interrelationships between organizational sector and various demographic or background variables available to us. We used crosstabulation analysis for this purpose and generated $\chi^2$ statistics to test for differences. Next we examined the zero order inter-correlation matrix for all of the scale variables in order of the sets described above. Finally, we tested for differences between government and voluntary sectors using Multiple Analysis of Variance or (MANOVA) with sets of scale variables as the dependent measures.

**RESULTS**

**Observed sample differences**

Cross-tabulation analysis revealed that the variable ‘sector’ is confounded by two demographic variables of note: First, sector was found to be related to organizational
identity. The questionnaire asked respondents to indicate their primary organizational identity from which we were able to categorize each respondent as predominantly having an evaluator identity (evaluation specialist, performance measurement specialist), or manager (program manager, program officer/staff person, senior manager, executive director). Respondents organizationally located in government were considerably more likely to self-identify as evaluators than was the case for those located in the not-for-profit sector: $\chi^2 (1, 201) = 51.42, p < .001$. This being the case, naturally respondents located in government tended to spend a higher proportion of their time doing evaluation than their counterparts in the voluntary sector, $\chi^2 (2, 240) = 31.16, p < .001$. Almost 45% of those respondents located in government indicated that they spent over 70% of their time working in evaluation while, conversely, over half of those located in the voluntary sector indicated that percentage of time spent on evaluation was less than 25%.

With the majority of respondents who self-identify as evaluators in government, we would expect that their knowledge about evaluation theory and practice would be higher as well. In fact, this turned out to the case: Survey participants located in the government sector self-reported significantly higher levels of knowledge than did their voluntary counterparts for both evaluation theory, $\chi^2 (3, 254) = 10.74, p < .01$ (52% vs. 33% reported ‘very good’ or ‘excellent’ level of knowledge of evaluation theory) and evaluation practice $\chi^2 (3, 254) = 21.27, p < .001$ (65% vs. 35% reported ‘very good’ or ‘excellent’ knowledge of evaluation practice).

We tested for several other differences between government and voluntary sector respondents and no statistically significant differences for gender, age group, education level or years of experience in their current function.

**Self reported perceptions about organizational evaluation capacity**

**Intercorrelations**

Table 2 shows the pattern of zero-order intercorrelations among all of our scale variables. In the Table we can see a high degree of intercorrelation among all of the variables representing various evaluation capacity constructs. In the Table the variables are ordered in sets corresponding to antecedent conditions, capacity to do evaluation and capacity to use it. We would expect, and for the most part found, higher intercorrelations within categories of constructs. We can also see a strong pattern of intercorrelation across categories of variables.

-- Insert Table 2 about here --

**Tests for Differences**

We used MANOVA to test for differences in self-reported measures of evaluation capacity. Given the observed confound between sector (government, voluntary) and organizational role (evaluator, manager), we conducted a 2 x 2 factorial between groups design. Separate tests were run on each of the three sets of constructs – *antecedents, capacity to do evaluation, and capacity to use evaluation* – each yielding *F* values for sector, organizational role, and the sector by organizational role interaction.
Figure 2 shows differences between sectors, Multivariate $F_{(4, 171)} = 5.63, p < .001$, and between organizational role, Multivariate $F_{(4, 171)} = 2.51, p = .04$ on antecedents conditions supporting evaluation capacity. No multivariate interaction effect was observed. With the exception of organizational support structures (formalization), voluntary sector respondents were more likely to rate antecedent variables significantly higher than their government counterparts. The converse was true for the formalization construct. By contrast, role differences were restricted to a single variable, organizational learning. Managers tended to rate this construct significantly higher than did evaluators.

No differences were observed with respect to the four capacity to do evaluation constructs. As Table 1 shows these variables – capacity to do, types of evaluation, stakeholder level and frequency of participation – were generally rated quite high (close to 5 on the 7 point Likert scale). We noted modest tendencies for voluntary sector participants to rate both stakeholder level and frequency of participation in evaluation higher than their government counterparts, but these differences were not strong enough to produce a multivariate effect.

Differences were noted, however, for the capacity to use evaluation constructs. As shown in Figure 3, differences emerged between sectors, Multivariate $F_{(4, 171)} = 4.77, p = .003$, but not between organizational role. No multivariate interaction effect was observed. No differences were observed for conditions mediating use but voluntary sector respondents were more likely to rate use of findings and use of process variables significantly higher than their government counterparts.

In summary, observed differences consistently suggest that voluntary sector respondents, almost uniformly respective of role, rated their organizations more favourably than did their government sector counterparts with respect to the antecedents or conditions supporting evaluation capacity (including organizational learning capacity and access to training), and the capacity to use evaluation whether by responding to evaluation findings or benefiting from evaluation by virtue of proximity to the process. A modest tendency in the voluntary sector toward greater stakeholder involvement in evaluation is consistent with the observed pattern of process use. Interestingly, these findings emerge in light of the observation that government sector internal evaluators reported tendencies to self-identify as evaluators, spend more time on evaluation, and have higher levels of knowledge about evaluation theory and practice, than did voluntary sector participants.

DISCUSSION

In empirical studies comparing group differences it is generally desirable to demonstrate group equivalency as a basis for comparison or to provide statistical controls to account for preexisting differences. In the present case, groups differed at the outset most notably in terms of government sector internal evaluators having primarily
evaluation responsibility and voluntary sector participants reporting management responsibility. While this may be an empirical finding of interest in its own right, it necessitated that we take organizational role into account in our analysis of sector differences.

Our 2 x 2 factorial multivariate analysis of variance yielded a very interesting pattern of results concerning self-reported opinions and perceptions about organizational evaluation capacity. First, all differences that we observed were between voluntary and government sector respondents. No effects for organizational role or interactions were observed. This finding reinforces the observation that despite the fact that voluntary sector respondents have management responsibility it may be said that they also tend to be the organization members responsible for internal evaluation. Given their role in management, then, voluntary sector respondents represent both supply and demand sides of evaluation; they are at the same time responsible for evaluation knowledge production as well as its use. Government participants in the main, on the other hand, were responsible merely for evaluation knowledge production.

Second, we found sector differences with respect to two sets of variables - contextual circumstances and capacity to use evaluation - but we failed to find differences with respect to the capacity to do evaluation. Let us examine each of these, one at a time. With respect to contextual conditions, voluntary sector respondents tended to perceive their organizations as being more capable of learning, more supportive of training and specifically, more supportive of evaluation capacity training. They also thought of their organizations as being less bureaucratic and rule-driven, which is unremarkable considering the respective structural considerations for government versus voluntary sector organizations (not to mention size). And given the relatively small size of voluntary sector organizations, it is not surprising that respondents perceived organizational learning capacity and perhaps commitment to professional learning and development to be greater for them than for government sector. Yet, a higher commitment to evaluation capacity building in the voluntary sector seems curious. It may be possible, as we have suggested above, that local (Compton et al., 2002) and systemic initiatives (Atkinson et al., 2005; United Way, 2005) to foster evaluative training is paying off in at least some areas of the voluntary sector. On the other hand, the perceived lack of such commitment by government participants is consistent with recent reports of a heightened need for professionalization and the development of advanced training opportunities in evaluation (Gussman, 2005; Hunt, 2006).

Despite a perceived differential commitment to training, learning and support for capacity building, differences were not apparent with respect to the capacity to do evaluation. These results are somewhat puzzling. Enhanced capacity building in the voluntary sector, as we suggested above, may be the result of a lack of capacity to do evaluation in the face of a growing need for it. Yet average scores on this variable were well above the midpoint on the Likert scale for both government and voluntary respondents. It may be argued that performance measurement has become an integral function in both sectors and therefore workers in this sector may have developed concomitant knowledge and skill that is relevant to evaluation. It seems likely also that, at least in government, such knowledge and skill would have derived mostly from practical experience, since, as Gussman (2005) observes, many beginning federal sector evaluators are more or less parachuted into the role or arrive there via otherwise indirect routes.
Equally puzzling is an absence of difference between the two groups in terms of the level and extent of involvement of stakeholders in evaluation; both groups reported fairly high levels of involvement. There was a tendency toward a perception of higher level and more frequent stakeholder involvement on behalf of voluntary sector respondents but this effect turned out not to be statistically significant. The range of stakeholders included in the respective measures was diverse and included program managers, developers and staff responsible for program implementation, as well as intended beneficiaries and sponsors/funders. It may be that government respondents were cuing on their propensity to work with members of the program community, especially managers and implementers. Yet we know that stakeholder engagement may be limited in government evaluation settings in light of the tension to remain independent or neutral in the name of objectivity/credibility (Mohan & Sullivan, 2006; VanLandingham, 2006). On the other hand, in the Canadian context, concerns have been raised that the bulk of evaluations conducted in a given year are targeted to program managers rather than higher level decision makers (Breem & associates, 2005; Gussman, 2005). Such a tendency might explain stakeholder involvement to some extent.

While training and organizational support appear to favor voluntary sector respondents, and no differences were found between groups with respect to the capacity to do evaluation, our data showed that the voluntary sector participants reported a higher capacity to use evaluation. Despite no difference between groups with respect to conditions mediating use, participants from the voluntary sector tended to report higher levels of both the ‘use of findings’ and the ‘use of process’, than did their government counterparts. We would argue that this finding is related to the fact that internal evaluators in the voluntary sector tend to be managers who have broad responsibilities including program and organizational evaluation. As we mentioned above, they tend to wear both supply and demand side hats, operating as both producers of evaluative information as well as consumers of it. No doubt they play a role in the symbolic use of evaluation to leverage funding from various sources, an increasingly challenging assignment given the current context of shrinking resources. But voluntary sector managers’ interest in using evaluation extends well beyond symbolic or persuasive use. According to findings from Hall et al. (2003) they also use evaluation findings for program and service modification. To that end, their valuing of the learning function of evaluation is underscored.

That program managers have responsibility for evaluation naturally contributes to greater perceived process use of evaluation as well. By virtue of their direct proximity to evaluation program managers would undoubtedly be developing knowledge and skill in evaluation logic and inquiry. We would argue that such knowledge and skill would spill into other aspects of their multi-faceted role (Cousins, 2007; Patton, 1997). On the other hand, even if program evaluators in our government sample are internal evaluators, at some level they remain external to the programs they evaluate. This being the case, program managers would be relatively more distanced from evaluations and therefore further from the potential benefits arising from the process of doing evaluation.

We conclude that despite similar capacities to carry out the evaluation function, voluntary sector consumers use evaluation to a greater extent. This finding is consistent with our working hypothesis that data use leads to data valuing. That voluntary sector managers necessarily must become directly involved in evaluation knowledge production
is a trigger or stimulus for experiencing use of evaluation. From our prior research in the education sector, (Cousins, Goh & Clark, 2005), we observed that if school-based decision makers successfully experience the use of evaluation they will tend to embrace it as a mode of knowing and understanding and will be more likely to engage in or otherwise require evaluation. Such an incentive may not be as apparent in government.

Other researchers have been giving direct consideration as to how to promote evaluation use in the federal government sector (e.g., Mohan & Sullivan, 2006; Perry et al., 2006). Mohan and Sullivan acknowledge the tension between the need to be responsive to user information needs and the need for independence. They suggest that evaluators ought to maximize both: responsiveness by consulting extensively, considering context, identifying and understanding stakeholder relationships, ensuring feasibility, responding to information needs, and using standards of practice to guide work. Maximizing independence would be come about by using methods that are accurate, reporting in a balanced way, being up front about politics; and differentiating between understanding beneficiary views and perspectives and being seen as championing / advocating them. Impartiality, they say, is central. We would add that understanding and navigating the organizational political landscape is essential. Evaluation after leans toward transparency and may therefore add credibility to programs on the one hand, while invoking organizationally defensive routines on the other.

In a related argument, evaluators should educate managers and policy makers about the beneficial relationship between performance management and evaluation as part of a comprehensive accountability system (Mohan et al., 2006). The suggestion is in accord with the proactive posture advocated by Young (2007) and Mayne (2007) with regard to evaluator’s role in engaging with performance measurement and accountability systems thereby heightening evaluation visibility and valuing. It will not be profitable for evaluators to work in silos or to maintain a mentality of knowledge production with no follow up.

All of this is good advice, we think. Yet none of these strategies are directly geared to using evaluation. Our hunch is that belief will follow practice; that successful use of evaluation (findings, process) will naturally lead to a heightened valuing of it. The problem to be solved then is how to get senior decision makers to experience the successful use of evaluation?

Perry et al. (2006) support the argument. They provide a case example showing that it is possible for a performance audit unit to ensure, (through stakeholder inclusion), that its work is relevant and has immediate, practical application while maintaining its role in providing independent oversight. According to the authors, “Far from raising concerns about our objectivity, this inclusiveness and transparency added a measure of credibility that increased the project’s impact and helped ensure our tools were used”. (p. 71). Immediate practical application implies use, and use implies increasing propensity to use again.

Mohan et al. (2007) propose a strategy that may work towards engaging senior level managers in evaluative thinking. They recommend that evaluators should educate managers and policy makers about the beneficial relationship between performance measurement and evaluation as part of a comprehensive accountability system. Evaluators can elicit their input, encouraging them to use performance measurement for making informed policy decisions. Admittedly there is a distance between
encouragement to use and actual use, yet strategies of this sort, we think, are on the right track.

Our study is timely in Canada given the recent endorsement of the Federal Accountability Act and its implications for the evaluation function (Government of Canada, 2007). Some would argue that more and better use of evaluation should be made at the systemic (government-wide) level rather than at the departmental level (e.g., McDavid & Huse, 2006) and evaluator independence is critical to fostering such use. In line with other colleagues (Mohan & Sullivan, 2006; Perry et al., 2006) we would challenge that assertion on the grounds that relationship building between evaluation knowledge producers and consumers is essential to use (Patton, 1997). Mayne and Rist (2006) favor an ‘arms-length’ relationship between program evaluators and managers, but a relationship nonetheless. Will it be enough? Young (2006), on the other hand, and in line with Mohan and Sullivan (2006), suggests that evaluators need to be actively involved in promoting and facilitating the function through developing networks and taking advantage of available opportunities to link evaluation to ongoing decision making.

Accountability in an era of new public management is very much about demonstrating program impact and results and value for money. Accountability in the voluntary sector is, at some level, a matter of survival; without evidence of program implementation and productivity, hopes of ongoing financial support are diminished. Yet unlike other sectors, accountability may not hinge on demonstrating program impact as much as it does on demonstrating learning. We observe that few resources tend to be available for evaluation in voluntary sector and therefore managers have little choice but to become directly involved in evaluation. Yet levels of self-reported evaluation use were comparatively high. A manager with responsibility for the evaluation function is well positioned to experience successful use of evaluation. Some would say ‘what gets measured gets done’ (e.g., Patton, 1997). Might we add to that, ‘what gets used gets done again’, the ‘what’ in this case being evaluation?

REFERENCES


### Table 1: Variable Construction and Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Alpha</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
</table>
| **Organizational Learning Capacity** (OLC) | **Perceived organizational learning capacity** including experimentation, teamwork, leadership, clarity of mission and transfer of knowledge. Composite (average) of 18 items on 7 pt. agree-disagree scale.  
- The organization's mission statement identifies values to which all staff must conform.  
- We can often bring new ideas into the organization.  
- We have opportunities for self-assessment with respect to goal attainment.  
- Most problem-solving groups include staff from a variety of functional areas or divisions.  
- New ideas from staff are treated seriously by management. | .958  | 4.48  | 1.21  | 306 |
| **Organizational Support Structures** | **Perceived extent to which organizational supports are present.** Includes 2 scales, one which focuses on training/professional development and the other which on the perceived degree of formalization in the organization:  
- **Organizational SS–Training:** Composite (average) of 7 items on a 7 pt agree-disagree scale.  
  - Staff training is relevant to our work  
  - The skills training that we receive can be applied to improve our work  
  - Learning that improves the work skills and knowledge of staff is encouraged  
  - Staff in this organization are encouraged to continuously upgrade and increase their knowledge and education levels  
  - Staff in this organization is provided with work-related skills training. | .929  | 5.09  | 1.21  | 314 |
|                                   | **Organizational SS – Formalization:** Composite (average) of 8 items on 7 pt. agree-disagree scale. NOTE: This scale has been shown to be negatively correlated to OLC (i.e., a LOWER score is actually preferred).  
- Information and decision-making must always go through proper channels  
- Most of our work must adhere to formal rules and procedures  
- Our work is usually closely monitored and inspected by management  
- Standard operating procedures have been established for almost every situation  
- We require approval in writing for the introduction of new work activities | .684  | 4.35  | 1.04  | 320 |
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Alpha</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Capacity to DO Evaluation                     | **Perceived capacity of the organization to conduct evaluation.** This includes factors such as resource availability, tools/resources, evaluation competencies. Composite (average) of 12 items, on 7 pt. agree-disagree scale.  
  - We have formal requirements to report on performance.  
  - Our organization provides us with the basic tools/resources to support evaluation (e.g. computers, software, copying, administrative support).  
  - Performance measurement is integral to our organizational accountability framework.  
  - Our organization has the knowledge and skills to oversee evaluations performed by external evaluators.  
  - Our organization provides positive encouragement to conduct evaluation.                                                                 | .907  | 4.83 | 1.2 | 314|
| Specific TYPES of Evaluation Activities       | **Perceptions about the types of evaluation activities within the organization** including formal planning or conducting program evaluations, performance measurement, or monitoring. Composite (average) of 17 items, on 7 pt. frequency scale.  
  - Produce reports for Boards of Directors and/or senior management  
  - Produce reports about program activities  
  - Review program documentation (e.g., participant records, case notes)  
  - Conduct formal program evaluations  
  - Assess the degree to which program goals/objectives are met | .920  | 4.85 | 1.17 | 303|
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Alpha</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder Participation in Evaluation</td>
<td>Perceived extent to which members of different stakeholder groups are involved in evaluation. Includes 2 scales, one which focuses on the level of participation and one which focuses on the frequency of participation: <strong>Level of Participation:</strong> Composite (average) of 6 items, on 7 pt. scale (1-low to 7-high). Includes stakeholder groups such as:  1. Staff responsible for implementing the program  2. Program managers or directors  3. Program developers  4. Intended beneficiaries of the program  5. Program sponsors or funders  <strong>Frequency of Participation:</strong> Same as above, but scale describes the frequency of participation (1 - never to 7 -always).</td>
<td>.813</td>
<td>4.55</td>
<td>1.27</td>
<td>294</td>
</tr>
<tr>
<td>USE of Evaluation Findings</td>
<td>Perceptions about how evaluation findings are used including decision-making, planning, and program improvements. Composite (average) of 10 items, on a 7 pt. frequency scale.</td>
<td>.908</td>
<td>5.0</td>
<td>1.21</td>
<td>315</td>
</tr>
<tr>
<td>USE of Evaluation Process</td>
<td>Perceptions about process use including learning about evaluation methods, (logic, skills, technical skills), increasing ownership and commitment, and integrating evaluation into work practices. Composite (average) of 17 items, on 7 pt. frequency scale.</td>
<td>.970</td>
<td>4.76</td>
<td>1.31</td>
<td>299</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Alpha</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Conditions Mediating Evaluation Use</td>
<td>Perceptions about the factors and conditions which may influence the use of findings including timeliness, quality, objectivity. Composite (average) of 19 items, on 7 pt. frequency scale.</td>
<td>.960</td>
<td>4.87</td>
<td>1.12</td>
<td>287</td>
</tr>
<tr>
<td>Variable</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1. Organizational learning</td>
<td>--</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2. Organizational Support Structures - Training</td>
<td>.713(**)</td>
<td>--</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3. Organizational Support Structures - Formalization</td>
<td>-.005</td>
<td>.217(**)</td>
<td>--</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4. Index for ECB (Sources of KS)</td>
<td>.578(**)</td>
<td>.593(**)</td>
<td>.081</td>
<td>--</td>
<td>.</td>
</tr>
<tr>
<td>5. Capacity to Do</td>
<td>.608(**)</td>
<td>.615(**)</td>
<td>.268(**)</td>
<td>.695(**)</td>
<td>--</td>
</tr>
<tr>
<td>6. Types of Evaluation Activities</td>
<td>.426(**)</td>
<td>.419(**)</td>
<td>.273(**)</td>
<td>.568(**)</td>
<td>.668(**)</td>
</tr>
<tr>
<td>7. Stakeholder LEVEL of Participation</td>
<td>.371(**)</td>
<td>.301(**)</td>
<td>.059</td>
<td>.355(**)</td>
<td>.457(**)</td>
</tr>
<tr>
<td>8. Stakeholder Frequency of Participation</td>
<td>.307(**)</td>
<td>.243(**)</td>
<td>.055</td>
<td>.310(**)</td>
<td>.384(**)</td>
</tr>
<tr>
<td>9. Conditions Mediating Evaluation Use</td>
<td>.494(**)</td>
<td>.475(**)</td>
<td>.190(**)</td>
<td>.572(**)</td>
<td>.724(**)</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).
Figure 1: Evaluation and organizational capacity
Figure 2: Mean ratings of antecedent variables by sector and organizational role
Panel A: Conditions Mediating Use

Panel B: Use of Evaluation Findings

Panel C: Use of Evaluation Process

Figure 3: Mean ratings of capacity to use evaluation variables by sector and organizational role