Assessing the Quality of Aboriginal Program Evaluations

Steve Jacob
Geoffroy Desautels
Laval University

Abstract: *Evaluations have gained in popularity in Canada since the 1990s, but statistical data indicate that the resources allocated to this management tool have not increased accordingly, despite the increased demand. During the same period, regardless of significant efforts to optimize governance, the Canadian federal government's management of issues related to Aboriginal peoples presents some weaknesses. Because evaluation may directly affect the administration of public programs, this study proposes a meta-evaluation of First Nations program evaluations. To do so, we replicate a methodology previously used by the Treasury Board Secretariat in 2004 to complete a vast study assessing the quality of evaluation in Canada. This article, based on the systematic analysis of a nonprobability sampling of more than 20 program evaluation reports, has applied the TBS's meta-evaluation techniques to the Aboriginal context. The results show that the evaluation of Aboriginal programs is of good, and even excellent, quality and suggest that the TBS's evaluation policy has had a definitive impact on evaluation quality.*

Keywords: *Aboriginal evaluation, evaluation standards, evaluative norms, meta-evaluation, quality of evaluation*

Résumé : *Depuis les années 1990, au Canada on observe un recours de plus en plus fréquent à l'évaluation. Toutefois, les statistiques démontrent que les ressources allouées à la pratique n'ont pas nécessairement suivi l'augmentation de la demande. Au cours de la même période, bien que d'importants efforts ont été déployés afin d'optimiser la gouvernance, des faiblesses demeurent au sein des processus de gestion du gouvernement fédéral canadien quant aux enjeux reliés aux peuples autochtones. Sachant que l'évaluation peut influencer l'administration des programmes publics, cette étude propose une méta-évaluation de l'évaluation de programme autochtone. Pour ce faire, on a utilisé une méthodologie précédemment utilisée par le Secrétariat du Conseil du Trésor en 2004, lors d'une vaste étude visant à évaluer la qualité de l'évaluation au Canada. Cet article, basé sur l'analyse systématique d'un échantillonnage non probabiliste de plus de 20 rapports d'évaluation des programmes, examine l'application des techniques de la méta-évaluation du SCT en contexte autochtone.*

Corresponding author: Steve Jacob, Faculty of Social Sciences, Department of Political Science, Laval University, Charles-de Koninck Pavillon, Office 4443, 1030 avenue des Sciences Humaines, Quebec, QC, Canada, G1V 0A6; steve.jacob@pol.ulaval.ca

Les résultats obtenus démontrent une évaluation de programmes autochtones de bonne qualité, même d'excellente qualité, et permettent de conclure que la politique d'évaluation du SCT donne des résultats concrets.

Mots clés : *évaluation autochtone, normes d'évaluation, normes évaluatives, méta-évaluation, qualité de l'évaluation*

Since the early 1990s, new governance processes based on results and accountability have driven governing bodies to justify their decisions to demonstrate the efficacy with which they spend public money (Dwivedi & Gow, 1999). Because of this, program evaluation has become increasingly popular within the Canadian public administration, notably with the introduction of the Canadian Evaluation Policy, drafted by the Treasury Board of Canada Secretariat (TBS, 2004a). The TBS's Evaluation Policy, whose latest version was adopted in 2009, is part of this new administrative paradigm pushing governments to better equip themselves to take more efficient and better informed decisions (TBS, 2009). Despite the fact that evaluations have gained in popularity in Canada since the 1990s (Birch & Jacob, 2005), statistical data indicate that resources allocated to this management field have not increased at the same rate as demand, making it more and more challenging for evaluators (TBS, 2003). Because the introduction of the policy has had a direct impact on evaluation use, the production of high quality evaluations is essential to foster efficient decision-making based on rigorous, valid, reliable, and credible recommendations and conclusions (Daigneault, 2010). In fact, because evaluations are more likely to be integrated into the administrative process, many researchers have paid closer attention to evaluation quality and, more concretely, to the negative impact of poor quality evaluations (Forss & Carlsson, 1997; Smith, 1999, quoted by Schwartz & Mayne, 2005). These studies have brought to light how overall quality can vary from one evaluation to the next. Because evaluation can impact the decision-making process, a poor quality evaluation can lead to erroneous decisions based on unreliable data, in areas such as fund allocation (Heinrich, 2012). At the same time, researchers have demonstrated the corollary of this reality, suggesting that as the quality of an evaluation increases, the more this evaluation is likely to be used by public administration decision-makers (Chelimsky, 1977; Christie & Alkin, 1999; Cousins & Leithwood, 1986; Schwartz & Mayne, 2005; Weiss & Bucuvalas, 1980).

The criteria used to evaluate quality are relatively well established. Several evaluation organizations have developed evaluation norms to guide evaluation practice and guarantee its overall quality (Australasian Evaluation Society, 2000; African Evaluation Association, 2006; DEGEVAL, 2008; Jacob & Boisvert, 2010; Société suisse d'évaluation, 2000). These norms are most frequently adapted from those of the American Joint Committee on Standards for Educational Evaluation (JCSEE) first published in 1981 and subsequently revised (JCSEE, 1994; Mbaïreweye & Jacob, 2012; Widmer, Landert, & Bachman, 2000). Following in

the JCSEE's footsteps, the Canadian standards are divided into four categories: utility, feasibility, propriety, and accuracy (Canadian Evaluation Society, 2010).

The introduction of standards monitoring evaluation practice promotes a more critical view of the quality of evaluations. This process of ex-post revision, also known as meta-evaluation, is frequently recommended by the evaluative norms of different evaluation organizations. Questions regarding validation and revision of evaluations have long been debated and studied in the literature (Cook & Gruder, 1978; Cooksy & Caracelli, 2005; Mayne, 2005; Patel, 2002; Schwartz & Mayne, 2005; Scriven, 2005; Stufflebeam, 2001, 2011; Worthen, 2001). The meta-evaluation process is a carbon copy of the evaluation process, with one main difference: its objective. Instead of looking into the value of a program, it rather seeks to validate the quality of the evaluation itself (Cooksy & Caracelli, 2005). Though evaluation standards are recent, meta-evaluation has been in use for quite some time. Its use increased progressively as more and more authors questioned evaluation quality (Forss & Carlsson, 1997; Guba, 1969), and it eventually became a necessary step to ensure the profession's credibility (Scriven, 2005; Stufflebeam, 2011).

Constantly putting the credibility of the evaluation process to the test is crucial if one wishes to integrate evaluation into the decision-making process. In fact, a TBS study (2006) about the government's role in evaluation quality mentions the challenges faced in the field of evaluation:

> A review of the literature on the role of evaluation in government found evaluation to be recognized as an important and longstanding function in Canadian government. However, the extent that evaluation has actually [been] integrated into government decision-making has been limited. (p. 1)

Jacob (2006) also points out that the institutionalization of Canadian evaluation seeks to ensure better quality and usefulness in the activities undertaken. Thus, quality and utility are essential conditions and are ultimately linked when evaluation procedures are to be undertaken on a regular basis. From a standpoint of resource optimization, while we hope to maximize the utility of public funds, the issue of quality naturally becomes central to the evaluation process.

RESEARCH OBJECTIVES

As stated earlier, evaluation quality can be directly linked to efficient decision-making and to sound administration. Yet, in Canada, despite the fact that important efforts have been implemented to optimize governance, one notices that governance issues relating to Aboriginal peoples present a series of weaknesses in the Canadian state's management processes. In fact, the management of Aboriginal issues is one of the strongest examples of ineffective governance in the country (Government of Canada, 2011). Unfortunately, this most alarming situation has persisted for years now, and recent studies do not indicate that it will improve in years to come. Nevertheless, over the last few years, several administrative initia-

tives have sought to foster the empowerment of indigenous communities. This empowerment strategy is related to the conclusions of the Royal Commission on Aboriginal Peoples, published in 1996, that urged the Canadian government to take greater consideration of First Nations, Métis, and Inuit traditional knowledge and encourage greater collaboration between its ministries when elaborating public policies, notably those related to Aboriginal health issues (Abele, 2006, p. 6). Despite some steps that have been taken toward Aboriginal self-government, thus encouraging the development of Aboriginal communities, there are persistent gaps, and statistical data still demonstrate important differences between Aboriginal peoples and the general Canadian population (Government of Canada, 2008).

Despite the fact that Aboriginal governments are undergoing profound changes and face many challenges, we do not necessarily expect that their program evaluation will be of poor quality, but we can anticipate that it will be different from traditional evaluation (Jacob & Desautels, 2013). As stated by Chouinard and Cousins (2007, p. 47), Aboriginal evaluation differs from traditional evaluation in that it shows greater attention to cultural context and places greater emphasis on the participation of stakeholders in a process of empowerment. Because First Nations, Métis, and Inuit peoples face so many socio-economic issues that hamper their communities' ability to flourish and develop, it seems natural to hope that maximizing social program performance and evaluation is prioritized. Realizing that evaluation is present in the decision-making processes more than ever, we would like to substantiate the quality of programs aimed at Aboriginals in Canada. Interest in Aboriginal program evaluation is of particular significance, as evaluation is the first factor that affects the utilization of recommendations and conclusions, which may lead to the betterment of public interventions (Weiss & Bucuvalas, 1980).

This study undertakes a meta-evaluation of First Nations program evaluations. A similar undertaking was completed by the TBS in 2004, which conducted a vast study aimed at assessing the quality of evaluation in Canada (TBS, 2004b). In its review of more than 100 program evaluations, this assessment highlighted a series of strengths and weaknesses related to issues in Canadian evaluation. From the data collected in their meta-evaluation, the research team was able to pull together a list of general criteria that can be used as an analytical chart, whose criteria are as follows: evaluation reports

- are clearly written, are concise and use simple language;
- clearly describe the program, policy, or initiative being evaluated, including its objectives, outputs, expected outcomes, reach and resources;
- have an assessment of the results achieved by the policy, program, or initiative;
- have a description of the evaluation, including its timing; the methodology; the evaluation objectives and issues; and how the evaluation fits into, and is important to, the overall operations of the department or agency;

- expose the limits of the evaluation, in terms of context, scope, methods, and conclusions;
- have an appropriate methodology (e.g., multiple lines of evidence);
- have conclusions that clearly address the main evaluation issues or relevance, success/impacts, and cost-effectiveness (depending on the type of evaluation—formative or summative);
- include only information necessary to understand findings, conclusions, and recommendations;
- present evidence-based and credible findings, for example:
  - evidence gathered in surveys of a representative group of participants, and compared to a comparable group of non-respondents
  - evidence derived from comparisons to baseline measures from the performance management system; and
  - qualitative evidence gathered from key informants who do not have a stake in the respective program or who are truly knowledgeable in the area of question;
- have conclusions and recommendations flowing logically from evaluation findings;
- have clear, attainable recommendations indicating actions to be taken and time frame; and
- provide analysis and explanation of exposure to risk of problems identified and in respect to recommendations made. (TBS, 2004b, p. 89)

In summary, the TBS study findings demonstrate that evaluation quality is relatively uneven within the Canadian federal government. Indeed, although most evaluations retained by the study were judged to be adequate, 23% were actually deemed inadequate. Despite this, the 2004 study establishes that Canadian evaluation procedures have improved over time. For example, the TBS study report maintains, in reference to post-2002 evaluations, that these were "addressing cost-effectiveness issues; methodological rigour; identifying alternatives; presentation of evidence-based findings; and, formal recommendations" (TBS, 2004b, p. 2). The report suggests that these tangible improvements in evaluation quality are explained by the efforts of the TBS to promote 2001's Evaluation Policy in different ministries and public organizations.

METHODOLOGY

The main goal of this study is to assess the quality of Aboriginal Public Programs evaluation reports. To achieve this, we used the evaluation model presented by the TBS in 2004 and applied it to a sample of Aboriginal program reports.

Data Sources and Samples

The review of the evaluation quality of First Nations, Métis, and Inuit programs was based on a nonprobabilistic sample of evaluation reports representative[1]

of Aboriginal programs' evaluation in Canada. In this study, any report was considered Aboriginal if it was produced through an evaluation process involving a public program in an Aboriginal community as defined by Indian and Northern Affairs Canada (INAC),[2] thus including Inuit communities in Nunavut, the Northwest Territories, the Yukon, Northern Québec (Nunavik), and Labrador; all First Nations, including those communities, towns, or villages not part of traditional territories or reservations (such as Winnipeg's Aboriginal community); and all Métis communities. To better define the characteristics of the populations targeted by the Aboriginal programs evaluation reports, the websites of INAC, the federal and provincial governments, and various private evaluation firms[3] were consulted. Although our initial plan was to bring together reports issued by different levels of government (national, provincial, and territorial), following an exhaustive search of various government websites and evaluation firms and a keyword search on Google it became obvious that, with over 60 reports published since 2003, INAC was by far the biggest player in the field in Canada.

Our goal was to draw up a comprehensive portrait of procedures linked to Aboriginal programs evaluation practices. To do so, we used a selection process similar to a stratified sampling, except that the criteria were applied qualitatively. A sample of 20 reports was deemed sufficient for our exploratory approach after we set aside reports that were judged irrelevant (evaluation type/date, programs covered by another evaluation, or programs similar to ones that were already selected). However, preliminary data and the preponderance of reports from INAC in our first sampling prompted us to increase the number of reports included to improve the final sample's representativeness. Figure 1 illustrates the selection process of the initial sample, and Figure 2 illustrates the second selection phase during which we tried to find additional relevant reports that had not been produced by or for INAC.

Initially, our objective was to classify various evaluation reports according to the level of government, publication date, and evaluation type (internal/external and formative/summative), but because of the lack of reports published by provincial and territorial governments, these criteria were only applied to reports published by INAC. The sample has some limitations, in that, because there is no centralized database of all Aboriginal evaluations available, it is based primarily on the results and impressions of our research on the websites of INAC, the federal and provincial governments, and various private evaluation firms. However, at first glance (Table 1), the sample seems to provide an adequate representation of program evaluations carried out in Aboriginal settings in Canada.

The sample comprises 27 Aboriginal program evaluation reports (see Appendix). Despite our efforts to diversify evaluation types, most reports were produced by a federal department: INAC. It is noticeable that INAC frequently involves outside collaborators in the evaluation process, which explains why more than a third of all evaluation reports are classified as using both internal and external

evaluation resources. Another interesting fact about our sample is that most of the reports are summative evaluations (74%). Lastly, most reports (85%) were produced after 2005.[4]

Analytical Tools

Analysis of the reports was completed by coding their contents using a two-sided analytical model. The first section allowed for the description of the evaluation reports' general traits:

1. Year of publication
2. Internal or external evaluation
3. Summative or formative evaluation
4. The agency or department commissioning the report
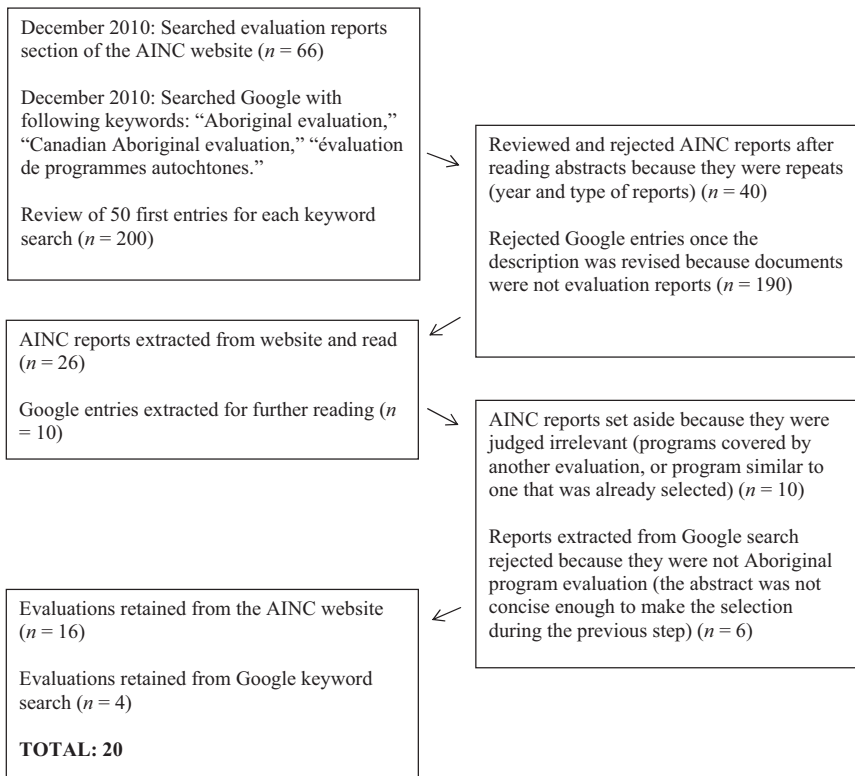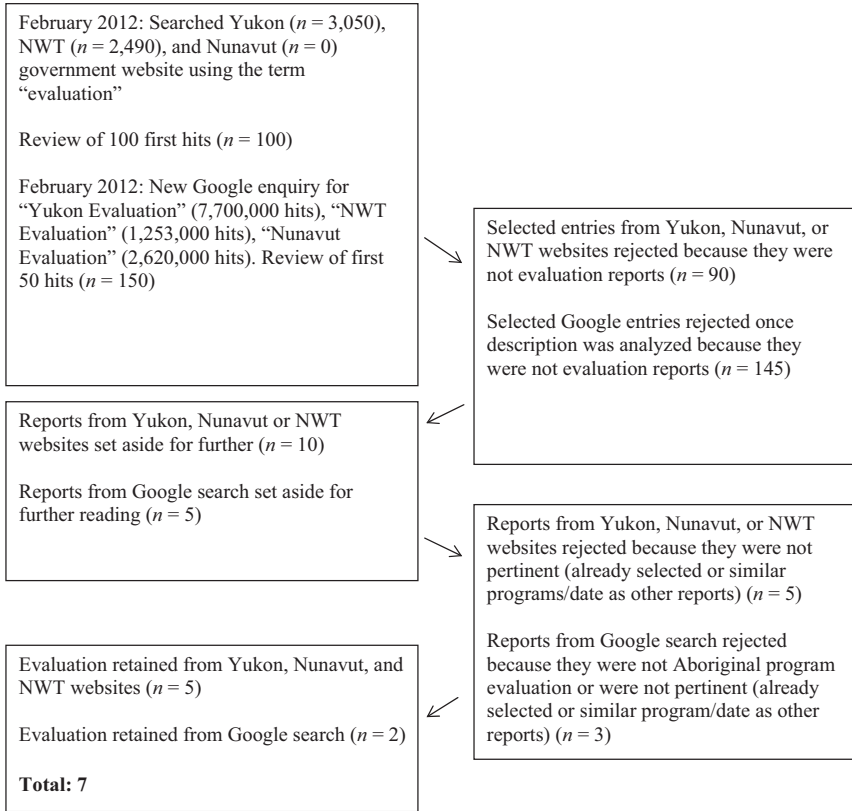5. The level of government involved (federal, provincial, territorial)

December 2010: Searched evaluation reports section of the AINC website ($n = 66$)

December 2010: Searched Google with following keywords: "Aboriginal evaluation," "Canadian Aboriginal evaluation," "évaluation de programmes autochtones."

Review of 50 first entries for each keyword search ($n = 200$)

Reviewed and rejected AINC reports after reading abstracts because they were repeats (year and type of reports) ($n = 40$)

Rejected Google entries once the description was revised because documents were not evaluation reports ($n = 190$)

AINC reports extracted from website and read ($n = 26$)

Google entries extracted for further reading ($n = 10$)

AINC reports set aside because they were judged irrelevant (programs covered by another evaluation, or program similar to one that was already selected) ($n = 10$)

Reports extracted from Google search rejected because they were not Aboriginal program evaluation (the abstract was not concise enough to make the selection during the previous step) ($n = 6$)

Evaluations retained from the AINC website ($n = 16$)

Evaluations retained from Google keyword search ($n = 4$)

TOTAL: 20

Figure 1. Sampling Process Phase 1

February 2012: Searched Yukon (*n* = 3,050), NWT (*n* = 2,490), and Nunavut (*n* = 0) government website using the term "evaluation"

Review of 100 first hits (*n* = 100)

February 2012: New Google enquiry for "Yukon Evaluation" (7,700,000 hits), "NWT Evaluation" (1,253,000 hits), "Nunavut Evaluation" (2,620,000 hits). Review of first 50 hits (*n* = 150)

Selected entries from Yukon, Nunavut, or NWT websites rejected because they were not evaluation reports (*n* = 90)

Selected Google entries rejected once description was analyzed because they were not evaluation reports (*n* = 145)

Reports from Yukon, Nunavut or NWT websites set aside for further (*n* = 10)

Reports from Google search set aside for further reading (*n* = 5)

Reports from Yukon, Nunavut, or NWT websites rejected because they were not pertinent (already selected or similar programs/date as other reports) (*n* = 5)

Reports from Google search rejected because they were not Aboriginal program evaluation or were not pertinent (already selected or similar program/date as other reports) (*n* = 3)

Evaluation retained from Yukon, Nunavut, and NWT websites (*n* = 5)

Evaluation retained from Google search (*n* = 2)

Total: 7

Figure 2. Sampling Process Phase 2

This section of the model was primarily used to provide a detailed description of the selected evaluation reports. The second section was based on the model developed by the TBS in its 2004 study. The TBS analytical model combines a series of documents pertaining to performance evaluation and quality assessment: (a) "Guide for the Review of Evaluation Reports," prepared by the Centre of Excellence for Evaluation, TBS, January 2004; (b) "Checklist Form for Internal Control of Evaluation Study: Deliverables/Report, Processes and Contractors' Work," prepared by Program Evaluation, HRDC, September 2003; (c) "Health Canada Evaluation Report Assessment Guide," prepared by the Departmental Program Evaluation Division, Health Canada, April 2003; (d) a framework for assessing the quality of evaluations, prepared by an external consultant for use by the Office of the Auditor General (but not implemented); and (e) excerpts from the OAG 1993 Report on Program Evaluation ("Criticisms re Evaluation Reports"), prepared by CEE (TBS, 2004b p. 7). From these documents used to assess the quality of evaluation reports, the Treasury Board of Canada elaborated its own analytical matrix. Its grid is divided into sections and subsections, allowing for a precise analysis of

Table 1. Characteristics of Sample Reports

Criteria		Number of reports	%
Government level	Federal	18	67
	Nunavut	1	4
	Northwest Territories	5	19
	Yukon	3	11
	Total	27	100
Evaluation type (1)	Internal	4	15
	External	15	55
	Internal and External	8	30
	Total	27	100
Evaluation type (2)	Formative	7	26
	Summative	20	74
	Total	27	100
Publication date	Prior to 2005	4	15
	From 2005 to 2009	11	41
	Since 2010	12	44
	Total	27	100

each individual report. The following list details the 10 sections, with a detailed explanation of some of the criteria we used.[5] For each criterion, except when we only coded "met"/"not met," we used a rating scale ranging from *poor* to *excellent*. We present this rating scale for three criteria to enable the reader to understand how the reports were evaluated.

1. Executive Summary
   1.1. Clearly and concisely written, coherent as a stand-alone document
      - **Poor**: Very poorly written or incomplete executive summary
      - **Below** average: Poorly written and omitting most elements of the evaluation
      - **Average**: Well written and the majority of the evaluation elements are presented
      - **Good**: Very well written and most of the evaluation elements are presented
      - **Excellent**: Very well written and all the elements of the evaluation are presented

    1.2. Presents key evaluation issues and answers these issues with relevant information through sound analysis

2.   Introduction and Context

    2.1. Describes the program, policy, or initiative being evaluated

- **Poor**: The program/policy/initiative is described briefly, but no details are given about objectives, funding, timetable, key activities, input/output, and theoretical framework
- **Below average**: The program/policy/initiative is described briefly, but very few details are given about objectives, funding, timetable, key activities, input/output, and theoretical framework
- **Average**: The program/policy/initiative is well described, but some key information about objectives, funding, timetable, key activities, input/output, and theoretical framework is missing
- **Good**: The program/policy/initiative is well described and most of the key information about objectives, funding, timetable, key activities, input/output, and theoretical framework is present in the report
- **Excellent**: All the elements of the program/policy/initiative are presented and very well detailed (objectives, funding, timetable, key activities, input/output, and theoretical framework)

    2.2. Describes intended beneficiaries and stakeholders involved

    2.3. Describes the cause-and-effect linkages among inputs, activities, outputs, and outcomes, and external factors contributing to success or failure

    2.4. Discusses resource allocation to policy, program, or initiative areas

    2.5. Identifies the role of the evaluation and its importance/significance at the time it was conducted

    2.6. Describes the key evaluation issues and questions linked to the program, policy, or initiative

3.   Methodology

    3.1. Description of the Methodology/Design

- **Poor**: The evaluation report briefly explains the methodology, but there are no further explanations of the steps undertaken
- **Below average**: The evaluation report briefly explains the methodology, but some key elements are not presented and it is not possible to know all the steps undertaken in the evaluation course
- **Average**: Methodological guidelines are presented, but there are very few or no details concerning, sampling strategy, tools, data triangulation, etc.
- **Good**: Methodological guidelines are presented, and there are details concerning most of the elements (sampling, tools, data triangulation, etc.)

- **Excellent**: Methodological guidelines are very well presented, and all the details are available concerning, sampling strategy, tools, data triangulation, etc.

   3.2. Multiple Lines of Evidence

   3.3. The limitations and trade-offs of the methodologies, data sources, and data used in the evaluation are clearly articulated

   3.4. Rigour

4. Key Findings

   4.1. Relevance

   4.2. Success

   4.3. Cost-effectiveness

   4.4. Delivery/Implementation

   4.5. Evaluation Issues

   4.6. Evidence-based Findings

   4.7. Analysis

5. Key Conclusions

6. Recommendations

7. Management Response

8. Action Plan

9. General/Other

10. Overall Assessment

Throughout the data coding process, the coder had to extract from the report all the elements needed to answer all the questions pertaining to key evaluation criteria. In the end, the results were tabulated according to explicit criteria for all the categories mentioned in the preceding list and followed the same principles as those set out by the Treasury Board of Canada's 2004 study:

> Most of the elements assessed in the reviews were rated on a five-point scale ranging from 1 (*poor*) to 5 (*excellent*), with the mid-point 3 indicating *average*. For the analyses, the scale ratings were collapsed into the three following categories: 1–2 (*inadequate*), 3 (*adequate*) and 4–5 (*more than adequate*). (TBS, 2004b, p. 11)

The coding was done by a single individual but was validated by a second individual in close collaboration with the first coder. The second individual had the criteria grid and was able to validate the work done by the coder. To ensure the validity and faithfulness of results, we employed a three-step coding process. First, all selected reports were read once to familiarize the coder with their contents and style. The analytical grid was rigorously applied to analyze each evaluation report. A few weeks after the first coding procedure, the reports' data were analyzed a second time and compared to the results of our first attempt at coding. Due to the precision of the criteria, there was very little fluctuation over time. However, if gaps existed between the two codings, the procedure was undertaken a third time to validate the final result. The final result was then validated by the second individual. This test-retest validation method was designed to ensure data coding

through time. On average, 6.5 hours were required to complete the entire process for each selected report (4.5 hours for the first analysis, including the first reading, and 2 hours for the second).

RESULTS

This section presents the key findings and details the quality of Aboriginal programs' evaluation reports.

Program Descriptions

Table 2 illustrates the data relating to the quality of program descriptions in the evaluation reports. The elements evaluated were description of initiative, beneficiaries and stakeholders, budget allocation review, account of underlying hypothesis and external factors, presentation of logical model, and description of links between cause and effect.

Table 2. Programme Description

Criteria	Met Criteria (%)	Ratings		
		Inadequate (%)	Adequate (%)	More than adequate (%)
Description of program, policy, or initiative	100	0	33	67
Description of beneficiaries and stakeholders[a]	93	11	56	33
Review of funding allocations	85	4	70	26
Description of unintended outcomes and of external influences[b]	37	0	75	22
Presentation of a logic model	19	n/a[c]	n/a	n/a
Description of causal links	33	0	89	11

[a]All or some beneficiaries and stakeholders.
[b]All or some unintended outcomes and external influences.
[c]"n/a" signifies no rating was made to this criterion.

All the Aboriginal program evaluation reports selected in our sample have a section pertaining to the description of the initiative or public policy. The criteria are respected adequately in about one third of cases, and more than adequately in 67% of the evaluation reports. A vast majority of selected documents have a section describing beneficiaries and stakeholders (93%), and a review of resource allocations is described in 85% of cases. However, few reports make mention of the logical model used to evaluate the programs (19% of reports). The absence of an analytical model impacts the description of hypothesis and of causal links. Descriptions of unintended outcomes and external factors are presented in only 37% of evaluation reports, while causal links are described in 33% of the cases.

Context

When it comes to the circumstances surrounding the evaluation process, we chose to focus on the following standards to measure the background information provided by the reports: depiction of evaluation objective, explanation of evaluation timeline or timeliness, justification of significance of evaluation, and description of issues and elements directly touched by the evaluation. The elements relating to the circumstances surrounding the evaluation process were generally respected. Almost all evaluation reports offered a detailed description of evaluation objectives (96% of cases), and, if present, the description was adequate in 62% and more than adequate in 27% of reports. The results were just as satisfactory for reports providing descriptions of the issues and elements to be evaluated (93% of cases), this criterion being fulfilled adequately or more than adequately in 92% of cases. More than three quarters of the reports selected (77%) mentioned the evaluation's timeline (again, 92% of those reports fulfilled the requirement adequately or more than adequately), and a similar number of reports (70%) detailed the evaluation's significance (the criterion was met adequately or more than adequately in 89% of cases).

Issue Coverage

Concerning the issue coverage in the course of an evaluation, the results speak for themselves: when it comes to pertinence, cost-benefit analysis, and management practices, Aboriginal program evaluation reports demonstrate excellent coverage. Pertinence is discussed in 96% of reports, goal attainment in 93% of cases, profitability in 81% of reports, program implantation and execution in 78% of cases, and management practices in 89% of the reports analyzed.

Methodology

Table 3 presents data on characteristics related to evaluation methodology. Discussions related to evaluation methodology vary from one report to the next, and there is significant disparity depending on the reference criterion. Although a description of applied methodologies and concepts (appearing in 96% of cases),

Table 3. Methodology

Criteria	Met criteria (%)	Ratings		
		Inadequate (%)	Adequate (%)	More than adequate (%)
Description of methods and designs applied[a]	96	12	50	38
Description of elements		n/a[b]	n/a	n/a
Sample size	89	n/a	n/a	n/a
Sampling method	33	n/a	n/a	n/a
Links between methods and elements	33	n/a	n/a	n/a
Technical document referrals	56	n/a	n/a	n/a
Instruments	89	n/a	n/a	n/a
Suitable design	85	0	91	9

[a]Description or listing of elements.
[b]"n/a" signifies no rating was made to this criterion.

sample size (in 89%), instruments used (89%), and a reasonable evaluation design (85%) is present in most reports, other criteria, such as sample selection method (appearing in 33% of the reports analyzed), linkage of methods and issues (in 33%), and references to technical documentation (in 56%), are not frequently mentioned.

Data Sources

When it comes to data collection during evaluations of Aboriginal programs, it is important to note that 89% of the evaluation reports included multiple data sources to support their observations. At the same time, 85% of the evaluation reports in our sample provide a suitable balance between quantitative and qualitative data sources. Nevertheless, few reports include data relative to external perspectives, this criterion being respected by only 30% of reports analyzed. The same was observed for data ensuing from an ongoing monitoring of program performance (7% of sample reports). It is also important to note that stakeholder perspectives are adequately presented in 48% of sample reports and more than adequately presented in 40% of cases, which means that in most evaluation reports, the views of a majority or all of the stakeholders involved were heard.

Table 4. Relevance Findings

Criteria	Met criteria (%)	Ratings		
		Inadequate (%)	Adequate (%)	More than adequate (%)
Evidence to demonstrate actual need	93	8	72	20
Evidence to demonstrate responsiveness to need	93	4	72	26
Evidence to demonstrate continued relevance to government priorities	89	0	67	33
Evidence to demonstrate there is no duplication	37	10	60	30

Note. Only those reports that could be evaluated on this criterion were scored.

Program Relevance

Table 4 contains data on findings linked to program relevance. On this point, our sample of Aboriginal program evaluation reports performs well. In almost all cases (93%), data aiming to illustrate need or beneficiary receptive are presented in the evaluation reports. In the same vein, the program evaluation reports present figures that demonstrate the program's links to government priorities (89%). In fact, only one criterion in this section was rarely met (37% of reports): an indication of whether the program is duplicated elsewhere or is counteracting other government programs.

Measuring Program Success

When it comes to program success, most of the evaluation reports (93%) presented a detailed description of program results and funding allocation in view of its efficiency. The reviewing of other factors linked to success was also fairly well documented by our sample reports (85% of reports). In fact, when documented, it was done adequately or more than adequately in all cases. Unintended outcomes are mentioned in 55% of reports, while 44% of sample reports comment on other programs, initiatives, or policies that may have impacted program results. More than 1 out of every 3 evaluation reports (37%) take the impact of other programs or policies into account when assessing program results. Program incrementality is never truly assessed, though 15% of sample reports do mention it.

Other Aspects Linked to Conclusions

Concerning other aspects related to evaluation reports, the results vary greatly. On the one hand, issues and evaluation questions are handled adequately or more than adequately in most reports (67% and 26%, respectively). Also, most conclusions are based on available data and are logical interpretations of the results and analysis (adequately in 67% of cases and more than adequately in 33% of cases). Despite this, once the data are extracted from the report, we were able to determine that very few evaluation reports presented an appropriate analysis. In most cases, the analyses were not based on previously presented data or the relationship between data and analysis was not clearly demonstrated.

The main observations referring to the evaluation reports' conclusions show that 93% of those conclusions were based on assessment criteria or referred explicitly to specific details. Also, in most cases (74%), the report findings reflected on lessons learned throughout the process. The criterion aimed at verifying if objective conclusions are drawn based on implementation, execution, or management practices is adequately or more than adequately covered by 83% of our sample reports. Lastly, unbiased conclusions based on data, relevance, program successfulness, or profitability are adequately or more than adequately detailed in 89% of cases.

Recommendations

When it comes to recommendations, most of the evaluation reports in our sample offered recommendations that logically arose from observations, conclusions, and recommendations (93% of evaluation reports respected this criterion). Nevertheless, only 52% of the reports included a recommendation on the program's overall funding. Also, 93% of evaluation reports presented adequately or more than adequately detailed, practical recommendations, and the same percentage of our sample touched upon alternative scenarios that account for possible practical concerns.

Clarity and Other Aspects of Evaluation Reports

Data on evaluation report clarity and on other aspects of the reports are featured in Table 5. When it comes to clarity and other aspects of the evaluation reports, most of them are written in a clear manner (59% are adequate and 33% are more than adequate). In all cases, technical data are reasonably well presented (74% adequately and 26% more than adequately), and good use is made of charts and graphs (74% adequately and 26% more than adequately). Some reports (11%) do fail to provide good quality technical annexes or fail at proper organization for ease of consultation, while all other reports are very clear and well organized.

Table 5. Report Clarity and Other Aspects

Criteria	Inadequate (%)	Adequate (%)	More than adequate (%)
Clearly written evaluation report	7	59	33
Appropriate presentation of technical information	0	63	37
Technical appendices are of high quality	11	48	41
Data presented fairly	0	74	26
Effective use of tables and charts	0	74	26
Report is well-organized and easy to follow	11	56	33

Overall Quality of Evaluation Report

Once each evaluation report's data were coded, we qualitatively assessed the report's quality. This assessment took into account the overall quality of the evaluation report. In particular, we looked to see if the report was well-written and well-organized and whether all the evaluation objectives were met, as well as the way in which they were answered. Almost all Aboriginal programs' evaluation reports were of adequate (48%) or more than adequate (44%) quality. But a small number of reports (7%) were too flawed (no conclusions drawn/no recommendations; incomplete annexes; no structure) to be rated as adequate.

DISCUSSION

The study results help us get a better grasp of the key features of the Aboriginal program evaluation process. Overall, it can be concluded that the Aboriginal evaluation process is of good, or even excellent, quality. In fact, our qualitative analysis led us to conclude that 92% of the evaluations in our sample can be deemed adequate or more than adequate. Beyond this overall appreciation, it seems important to consider what features lead to quality Aboriginal program evaluation. Table 6 provides a synthesis of the strengths and weaknesses identified in our study sample.

These results give us a detailed portrayal of Aboriginal program evaluation; when we compared them to the main findings of the 2004 TBS study, which reviewed evaluation quality in federal agencies and departments, we were able to find some interesting similarities. First, we notice that most strengths identified by the 2004 TBS study were the same as those identified in our sample: a thorough

Table 6. Strengths and Weaknesses of Aboriginal Programs Evaluation Process

Strengths most commonly identified in Aboriginal program evaluation	Weaknesses most commonly identified in Aboriginal program evaluation
Includes a detailed description of program or initiative, including funding, beneficiaries, and stakeholders. Funding allocation is well presented.	Few reports include a logical model of the program and a minority of them describes underlying assumptions and external factors.
Issues relative to evaluation context are well presented, in particular evaluation goals and issues and questions.	Sampling methods and links between data and methods used are rarely presented.
Aboriginal program evaluations offer excellent coverage of assessment tools, particularly in relation to relevance, success, and profitability.	Use of ongoing performance monitoring data is not frequent.
Coverage of management practices is good.	Non-stakeholder opinions are rarely taken into account in Aboriginal program evaluation reports.
Method and design are described in all Aboriginal program evaluation reports. Sample size and research tools are also elements described in most evaluation reports.	
Most evaluation reports present multiple data sources and an appropriate mix of qualitative and quantitative approaches.	Fewer than half of reports provide data on duplication.
Reports discussion of program need to demonstrate the genuine need for the program, receptivity to needs, and to show the continued relevance of the program when it comes to government priorities.	Factors related to implementation are rarely discussed.
Most of evaluations describe results relative to program success. Also most evaluations consider other factors that may contribute to program success.	Analysis is inadequate in almost half of the reports.
Evaluation issues and questions are adequately discussed, and findings are based on data and flow logically from the interpretation of data and analysis.	
Generally, conclusions present other lessons learnt. Also, conclusions are based on explicit judgement criteria or benchmarks.	
Reports contain recommendations based on significant findings and flow logically from the interpretation of data and analysis.	
The reports achieve the highest standards for clarity and overall quality of presentation.	

description of the program or initiative under evaluation, including detailed accounts of available resources, stakeholders, and program beneficiaries; a clear statement of evaluation objectives; use of multiple data sources in evaluation design; well-presented findings, in particular, those related to relevance, execution, and implementation of formal recommendations or of proposed changes that logically arise from the evaluation's conclusions as well as well-structured and clearly written report findings. Though evaluation strengths were constant for both studies, it is not the case for all weaknesses outlined by the 2004 TBS study. Weaknesses in that study included superficial coverage of issues pertaining to profitability, missing data on methodological approaches used in evaluation, and use of data to demonstrate need and program reception. Our program evaluation sample included more recently published reports than those analyzed by the TBS in 2004. It is interesting to note that Aboriginal program evaluation is of excellent quality. Furthermore, this is particularly important in light of the fact that the TBS had remarked on the significant improvement in the quality of evaluation reports after 2002 compared to those published in previous years:

> A noticeable improvement on a number of criteria was observed, however, when we compared evaluations completed prior to April 2002 with those done after this point in time. The latter, more recent evaluations show a significant improvement in quality, suggesting that TBS's April 2001 Evaluation Policy may have had a favourable impact. (TBS, 2004b, p. 40)

In our sample, 88% of the reports were published after 2005 and 96% after 2002.

The goal of our study was to assess Aboriginal program evaluation quality, and we feel our findings are encouraging. Aboriginal peoples in Canada face a great many challenges when it comes to governance, but to know that decision-making processes can rely on program evaluations whose quality is good or excellent leads us to believe that initiatives such as the TBS evaluation policy can lead to improved conditions. It goes without saying that widespread systematic use of program evaluation is not yet integrated into administrative practice (TBS, 2006; Jacob, 2006) and that there is still room for improvement when evaluation is applied, but it is undeniable that better quality evaluations will stimulate further use (Christie & Alkin, 1999; Schwartz & Mayne, 2005; Weiss & Bucuvalas, 1980).

By assessing Aboriginal program evaluation, our study has allowed us to get a detailed overall portrayal of this process. Comparing the TBS study was interesting in the sense that it offered us the opportunity to see the evolution of quality, despite the fact that our sample was restricted to only one policy sector. Further comparative studies using a similar coding method could benefit from an updated version of the 2004 TBS study so that comparisons can be made between the quality of Aboriginal program evaluation and other federal public policy areas within a specific timeframe. Also, because the institutionalization of evaluation advances so slowly (Jacob, 2005), a study aimed at qualifying and better understanding how evaluation reports are used in the decision-making process can lead to a better understanding of the impact of evaluation quality on governance processes. Last,

it is important to mention that despite our best intentions, elements such as evaluation team composition, justification of methodological methods, and description of the evaluation process were not systematically present in every evaluation. The same observation was made by the Treasury Board of Canada when it conducted its own study on the quality of federal agencies' program evaluation (2004, cited by Daigneault, 2010, pp. 208–209). In forthcoming studies, one should therefore consider Daigneault's recommendation "to contact organizations and evaluation units, for instance by submitting study results, in order to gather missing information and validate analyses and judgments about their evaluation reports" (Daigneault, 2010, p. 208, trans.). A more complex study, integrating meetings and interviews with key players, could complement our research project, refining and contextualizing the findings of our documentary analysis.

ACKNOWLEDGEMENT

NOTES

1   The representativeness of the sample selection process is presented in Tables 1 and 2.

2   In 2011, Indian and Northern Affairs Canada was renamed and is now known as Aboriginal Affairs and Northern Development Canada.

3   We consulted the websites of eight independent evaluation firms that were referred by the Canadian Evaluation Society: Goss Gilroy, Science-Metrix, Raymond Chabot Grant-Thornton, Capra International, Universalia, Johnson Research, Cathesis Consulting, and Harry Cumming and Associates Inc.

4   The oldest evaluation reports date back to 2001.

5   Readers are encouraged to consult the TBS Study Annexe (TBS, 2004b, pp. 46–62)for a more detailed description of the complete analytical grid used in 2004.

REFERENCES

Abele, F. (2006). Between respect and control: Traditional Indigenous knowledge in Canadian public policy. In M. Orsini & M. Smith (Eds.), *Critical policy studies: Contemporary Canadian approaches* (pp. 233–255). Vancouver, BC: UBC Press.

Australasian Evaluation Society (AES). (2000. Code of ethics. Retrieved from http://www.aes.asn.au/about/

African Evaluation Association (AfrEA). (2006). Evaluation standards and guidelines. Retrieved from http://www.afrea.org/?q=fr/page/resources

Birch, L., & Jacob, S. (2005). Program evaluation in Canada seen through the articles published in CJPE. *Canadian Journal of Program Evaluation*, *20*(3), 1–16.

Canadian Evaluation Society. (2010). CES guidelines for ethical conduct. Retrieved from http://www.evaluationcanada.ca/site.cgi?s=5&ss=4&_lang=EN

Chelimsky, E. (1977). *A symposium on the use of evaluation by federal agencies* (Vol. 2). McLean, VA: Mitre Corporation.

Chouinard, J., & Cousins, J. B. (2007). Culturally competent evaluation in First Nations and Inuit communities: A review of literature. *Journal of Multidisciplinary Evaluation*, *4*(8), 40–57.

Christie, C. A., & Alkin, M. C. (1999). Further reflections on evaluation misutilization. *Studies in Educational Evaluation*, *25*(1), 1–10. http://dx.doi.org/10.1016/S0191-491X(99)00006-1

Cook, T. D., & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Review*, *2*(1), 5–51. http://dx.doi.org/10.1177/0193841X7800200101

Cooksy, L. J., & Caracelli, V. J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation*, *26*(1), 31–42. http://dx.doi.org/10.1177/1098214004273252

Cousins, B., & Leithwood, K. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, *56*(3), 331–364. http://dx.doi.org/10.3102/00346543056003331

Daigneault, P.-M. (2010). L'examen de qualité des évaluations fédérales : une méta-évaluation réussie? *Revue canadienne d'évaluation de programme, 23*(2), 191–224.

DEGEVAL. (2008. DeGEval-Standards. Gesellschaftfür Evaluation e.V. Retrieved from http://www.degeval.de/degeval-standards/

Dwivedi, O. P., & Gow, J. I. (1999). *From bureaucracy to public management: The administrative culture of the Government of Canada*. Toronto, ON: University of Toronto Press.

Forss, K., & Carlsson, J. (1997). The quest for quality—Or can evaluation findings be trusted? *Evaluation*, *3*(4), 481–501. http://dx.doi.org/10.1177/135638909700300407

Government of Canada. (2008). Healthy Canadians: A federal report on comparable health indicators. Ottawa, ON: Health Canada.

Government of Canada. (2011). Canada's financial performance in an international context. Department of Finance. Retrieved from http://www.fin.gc.ca/budget05/bp/bpa2-eng.asp

Guba, E. J. (1969). The failure of educational evaluation. *Educational Technology*, *9*, 29–38.

Heinrich, C. J. (2012). How credible is the evidence, and does it matter? An analysis of the Program Assessment Rating Tool. *Public Administration Review*, *72*(1), 123–134. http://dx.doi.org/10.1111/j.1540-6210.2011.02490.x

Jacob, S. (2005). La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques (France, Belgique, Suisse et Pays-Bas). *Revue francaise de science politique*, *55*(5–6), 835–864. http://dx.doi.org/10.3917/rfsp.555.0835

Jacob, S. (2006). Trente ans d'évaluation de programme au Canada: l'institutionnalisation interne en quête de qualité. *Revue française d'administration publique, 119*, 515–532.

Jacob, S., & Boisvert, Y. (2010). To be or not to be a profession: Pros, cons and challenges for evaluation. *Evaluation: The International Journal of Theory, Research and Practice*, *16*(4), 349–369.

Jacob, S., & Desautels, G. (2013). Evaluation of Aboriginal programs: What place is given to participation and cultural sensitivity? *International Indigenous Policy Journal*, *4*(2), 1–29.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.

Mayne, J. (2005). Ensuring quality for evaluation: Lessons from auditors. *Canadian Journal of Program Evaluation*, *20*(1), 37–64.

Mbaïrewaye, M., & Jacob, S. (2012). Cultural dimensions of the African evaluation guidelines. In J.-C. Barbier & P. Hawkins (Eds.), Evaluation cultures: Sense-making in complex times (pp. 205–237). New Brunswick, NJ: Transaction.

Patel, M. (2002). A meta-evaluation, or quality assessment, of the evaluations in this issue, based on the African Evaluation Guidelines: 2002. *Evaluation and Program Planning*, *25*(4), 329–332. http://dx.doi.org/10.1016/S0149-7189(02)00043-5

Schwartz, R., & Mayne, J. (2005). Assuring the quality of evaluative information: Theory and practice. *Evaluation and Program Planning*, *28*(1), 1–14. http://dx.doi.org/10.1016/j.evalprogplan.2004.10.001

Scriven, M. (2005). Metaevaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 53–59). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412950558.n340.

Société suisse d'évaluation. (2000). Standards SEVAL. Retrieved from http://www.seval.ch/fr/documents/SEVAL_Standards_2001_fr.pdf

Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, *22*(2), 183–209.

Stufflebeam, D. L. (2011). Meta-evaluation. *Journal of Multidisciplinary Evaluation*, *7*(15), 99–158.

Treasury Board of Canada Secretariat (TBS). (2003). Evaluation policy: Results-based management and accountability framework (RMAF). Retrieved from https://www.tbs-sct.gc.ca/cee/tools-outils/polrmaf-polcgrr-eng.asp

Treasury Board of Canada Secretariat (TBS). (2004a). Guide for the review of evaluation reports. Ottawa, ON: Centre of Excellence for Evaluation. Retrieved from http://www.tbs-sct.gc.ca/cee/tools-outils/grer-gere-eng.asp

Treasury Board of Canada Secretariat (TBS). (2004b). Review of the quality of evaluations across departments and agencies. Retrieved from http://www.tbs-sct.gc.ca/cee/pubs/review-examen2004-eng.pdf

Treasury Board of Canada Secretariat (TBS). (2006). Roles for government in evaluation quality assurance: Discussion paper. Ottawa, ON: Centre of Excellence for Evaluation. Retrieved from http://www.tbs-sct.gc.ca/cee/tools-outils/qae-aqe/rgeqa-rgaqe-eng.asp

Treasury Board of Canada Secretariat (TBS). (2009).Policy on evaluation. Retrieved from http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024&section=text

Weiss, C. H., & Bucuvalas, M. J. (1980). Truth tests and utility tests: Decision-makers' frame of reference for social science research. *American Sociological Review*, *45*(2), 302–313. http://dx.doi.org/10.2307/2095127

Widmer, T., Landert, C., & Bachman, N. (2000). *Standards d'évaluation de la Société suisse d'évaluation (Standards SEVAL)*. Retrieved from http://www.seval.ch/fr/documents/SEVAL_Standards_2001_fr.pdf

Worthen, B. R. (2001). Whither evaluation? That all depends. *American Journal of Evaluation*, *22*(3), 409–418.

AUTHOR INFORMATION

Steve Jacob is a full professor in the Department of Political Science at Laval University. He is the founder and director of the research laboratory on public policy performance and evaluation (PerƒEval). He conducts research dealing with the mechanisms of performance management and evaluation: professionalization, institutionalization, and capacity building in Europe and Canada, ethics in evaluation, and participatory approaches.

Geoffroy Desautels is a professional researcher in the PerƒEval laboratory and a program evaluator at Laval University. He holds an MA in political science from Laval University and a BA in political science from the University of Montreal. His primary research interests are ethics in program evaluation and participatory evaluation.

APPENDIX. LIST OF PROGRAM EVALUATION REPORTS SAMPLED

Direction générale de l'évaluation de la mesure du rendement et de l'examen (DGEMRE). (2010). *Évaluation de la mise en œuvre du Fonds d'infrastructure pour les Premières nations (Numéro de projet: 08045)*. Affaires indiennes et du Nord Canada. (Report no. 16)

Comité de l'évaluation, de la mesure du rendement et de l'examen (CEMRE). (2010). *Évaluation sommative de la Contribution au titre de services de consultation fournis aux Inuits dans le Sud*. Affaires indiennes et du Nord Canada. (Report no. 3)

Consortium formé de Goss Gilroy Inc. et de Hollett and Sons Inc. (2011). *Évaluation de l'accord de subvention de la Première nation de Miawpukek*. Affaires indiennes et du Nord Canada. (Report no. 17)

Departmental Audit and Evaluation Branch and Goss Gilroy. (2005). *Evaluation of the Cultural/Education Centres Program*. Indian and Northern Affairs. (Report no. 23)

DGEMRE. (2008). *Évaluation des Programmes et de la Politique de gestion des sites contaminés d'AINC*. Affaires indiennes et du Nord Canada. (Report no. 18)

DGEMRE. (2010). *Évaluation des Subventions pour la sensibilisation à la protection contre les incendies dans les écoles administrées par les bandes et les écoles fédérales*. Affaires indiennes et du Nord Canada. (Report no. 14)

DGEMRE. (2009). *Évaluation du Programme d'information publique et de défense des intérêts*. Affaires indiennes et du Nord Canada. (Report no. 12)

DGEMRE et les consultants du Centre de Gestion Publique Inc. (2010). Le rapport a été rédigé par T.K. Gussman Associates Inc. *Évaluation sommative du program d'immobilisations et d'entretien*. Affaires indiennes et du Nord Canada. (Report no. 13)

DGEMRE et Donna Cona. (2010). *Évaluation de l'incidence des commissions des traités*. Affaires indiennes et du Nord Canada. (Report no. 1)

Direction générale de l'évaluation, de la mesure du rendement et de l'examen, Secteur de la vérification et de l'évaluation. (2010). *Évaluation formative du Programme d'enseignement primaire et secondaire dans les réserves*. Affaires indiennes et du Nord Canada. (Report no. 2)

Direction générale de l'évaluation et de la vérification interne avec l'aide de Rosemary Trehearne et associés. (2005). *Évaluation du Programme pour la prévention de la violence familiale chez les Premières Nations*. Affaires indiennes et du Nord Canada. (Report no. 5)

DPRA Canada en collaboration avec T.K. Gussman Associates. (2009). *Évaluation des initiatives de guérison communautaire offertes par la Fondation autochtone de guérison*. Affaires indiennes et du Nord Canada. (Report no. 9)

Engage Strategies. (2010). *Temporary Assignment Activity between Yukon First Nations and the Government of Yukon*. Workplace Diversity Employment Office Government of Yukon. (Report no. 7)

Evaluation Directorate Strategic Policy and Research Branch Human Resources and Skills Development Canada. (2009). *Summative Evaluation of the Employment Benefits and Support Measures Delivered Under the Canada-NWT Labour Market Development Agreement, NWT*. (Report no. 27)

Hanson, G., Lloyd, R., and Lorimier, B. (2004). *Evaluation of the Social Housing Program*. Yukon Housing Corporation and Canada Mortgage Housing Corporation. (Report no. 6)

Millennium Partners. (2002). *Building Nunavut Through Decentralization: Evaluation Report*. Department of Executive and Intergovernmental Affairs, Government of Nunavut. (Report no. 4)

J. Carey Consulting Evaluation Plus Ltd. (2010). *NWMOG Project Summative Evaluation*. Northwest Territories. (Report no. 25)

KPMG. (2001). *Evaluation of the P3 Pilot Initiative: Final Report*. Government of the Northwest Territories. (Report no. 8)

Secteur de la vérification et de l'évaluation. (2008). *Évaluation de la mise en œuvre de la Politique sur l'analyze comparative entre les sexes d'AINC*. Affaires indiennes et du Nord Canada. (Report no. 11)

PRA Inc. (2010). *Évaluation sommative des contributions aux bandes indiennes pour l'administration de l'inscription*. Affaires indiennes et du Nord Canada. (Report no. 15)

R. A. Malatest et Associates Ltd. (2011). *Evaluation of the Protection Against Family Violence Act*. Department of Justice NWT. (Report no. 26)

Ressources humaines et développement social du Canada. (2007). *Évaluation de l'Initiative de réinvestissement de la Prestation nationale pour enfants*. Affaires indiennes et du Nord Canada. (Report no. 19)

Société d'experts conseils non-identifiée. (2009). *Évaluation du program Premières nations sur Rescol*. Affaires indiennes et du Nord Canada. (Report no. 20)

Terriplan Consultants. (2009). *Territorial Medical Travel Program Northwest Territories*. Evaluation, Northwest Territories. (Report no. 24)

TNS Canada Facts et Harvey McCue Consulting. (2010). *Évaluation formative du Programme d'éducation postsecondaire*. Affaires indiennes et du Nord Canada. (Report no. 10)

Wildeor Wildlife Research and Consulting. (2008). *Evaluation of the Klondike Soft Gold Program*. Yukon Department of Environment. (Report no. 21)

Zapf, M. K. (2004). *An Evaluation of the Whole Child Project – Whitehorse Yukon. RCMP National Youth Strategy*. (Report no. 22)