



Treasury Board of Canada
Secrétariat

Secrétariat du Conseil du Trésor
du Canada

Canada

Evaluation of the 2009 Policy on Evaluation

Published: 2015-00-13

© Her Majesty the Queen in Right of Canada,
represented by the President of the Treasury Board, 2015

Published by Treasury Board of Canada, Secretariat
90 Elgin, Ottawa, Ontario, K1A 0R5, Canada

Catalogue Number: BT22-206/2015E-PDF
ISBN: 978-0-660-25703-7

This document is available on the Government of Canada website, Canada.ca

This document is available in alternative formats upon request.

Aussi offert en français sous le titre : Évaluation de la Politique sur l'évaluation de 2009



Evaluation of the 2009 Policy on Evaluation

Acknowledgements

This evaluation was conducted by the Centre of Excellence for Evaluation of the Treasury Board of Canada Secretariat and by an external consulting team composed of Natalie Kishchuk, CE, of Program Evaluation and Beyond Inc. and Benoît Gauthier, CE, of Circum Network Inc. The Centre of Excellence for Evaluation produced this final report, for which the external consultants provided a quality assurance review.

The Centre of Excellence for Evaluation wishes to thank the members of the advisory committees, who provided advice on the planning, conduct and reporting of the evaluation.

Table of Contents

Executive Summary

Background

Methodology

Findings and Conclusions

Performance of the Policy and Status of Policy Outcomes

Relevance and Impact of the Three Major Policy Requirements

Approaches Used to Measure Policy Performance

Other Findings

Conclusions

Recommendations

1.0 Introduction

1.1 Purpose of the Evaluation of the *Policy on Evaluation*

1.2 Background and Context

1.2.1 Evolution of the Federal *Policy on Evaluation* and the Context for Policy Renewal in 2009

1.2.2 The International Context for Evaluation

1.2.3 Introduction of the 2009 *Policy on Evaluation*

1.3 Overview of the Federal Evaluation Function

1.4 Implementation of the *Policy on Evaluation*

1.5 Context for Policy Renewal in 2014

2.0 Evaluation Approach and Design

2.1 Approach and Design

2.2 Methodology

2.3 Governance

2.4 Evaluation Period and Questions

2.5 Limitations

3.0 Findings

3.1 Performance of the Policy and Status of Policy Outcomes

3.1.1 Baseline Results for Policy Outcomes (evaluation questions 5 and 6)

3.1.2 Factors Influencing Outcome Achievement (evaluation question 8)

3.1.3 Sustainability of Outcomes (evaluation question 7)

3.2 Application of the Three Major Policy Requirements

3.2.1 Comprehensive Coverage (evaluation question 3)

3.2.2 The Five-year Frequency for Evaluations (evaluation question 2)

3.2.3 Core issues (evaluation question 1)

3.3 Approaches to Measuring Policy Performance (evaluation question 4)

3.4 Other Findings

4.0 Conclusions

5.0 Recommendations

Appendix A: Evolution of Evaluation in the Federal Government and the Context for Policy Renewal in 2009

Evolution of Policy Requirements for Extent and Frequency of Evaluation Coverage and Evaluation Issues

Context of the Federal Evaluation Function Leading Up To Policy Renewal in 2009

A 2006 Legislated Requirement for Evaluation

A Shift in Emphasis for Evaluation to Support Expenditure Management

Appendix B: Implementation Review of the 2009 *Policy on Evaluation*

Implementation of Policy Requirements Related to Leadership, Governance and Planning

Heads of Evaluation

Departmental Evaluation Committees

Departmental Evaluation Plans

Capacity for Fully Implementing Policy Requirements

Resource Allocation

Human Resources Capacity Building

Strategies Used to Implement Coverage Requirements

Performance Measurement to Support Evaluation

Implementation Progress in Small Departments and Agencies

Challenges of Policy Implementation

Appendix C: Purpose of the Evaluation of the 2009 *Policy on Evaluation*, Methodology and Governance Committees for the Evaluation

Purpose of the Evaluation

Methodology

Case Studies

Stakeholder Consultations

Surveys

Data Analysis

Process Mapping

Document Review

Literature Review

Governance Committees for the Evaluation of the 2009 *Policy on Evaluation*

Appendix D: Contribution Theory of the 2009 *Policy on Evaluation*, Generic Evaluation Process Map of the Life Cycle of a Departmental Evaluation, and Logic Model for Implementing the 2009 *Policy on Evaluation*

Abbreviations Used in Figures 7, 8 and 9

Endnotes

Executive Summary

Background

In 2013–14, the Treasury Board of Canada Secretariat conducted an evaluation of the 2009 [*Policy on Evaluation*](#). The evaluation assessed the performance of the policy; established a baseline of policy results—in particular, those related to evaluation use and utility; and identified opportunities to better support departments in meeting their evaluation needs through flexible application of the policy and the associated [directive](#) and [standard](#). This report documents the evaluation's key findings, conclusions and recommendations.

Methodology

The evaluation team was composed of external consultants as well as analysts from the Secretariat's Centre of Excellence for Evaluation. The external team members assessed policy performance, and the internal team members assessed policy application. The evaluation used both qualitative methods (case studies, process mapping, document and literature reviews, and stakeholder consultations with deputy heads, assistant deputy ministers, central agencies and others) and quantitative methods (analyses of monitoring data and surveys of program managers, evaluation managers and evaluators). The external and internal evaluation team members provided a challenge function for each other's work and assured the quality of evaluation products.

Findings and Conclusions

Performance of the Policy and Status of Policy Outcomes

Regarding the performance of the policy, the evaluation found the following:

1. In general, the evaluation needs of deputy heads and senior managers were well served under the 2009 *Policy on Evaluation*. Senior management was able to draw strategic insights to support higher-level decision making. At the same time, efforts to meet the policy's coverage requirements sometimes made evaluation units less able to respond to senior management's emerging needs.
2. The policy had an overall positive influence on meeting program managers' needs, and first-time evaluations of some programs were useful. However, program managers whose programs were evaluated as part of a broad program cluster or high-level Program Alignment Architecture entity sometimes found that their needs were not met as well as before 2009, when the program was evaluated on its own.
3. Central agencies found that evaluations were increasingly available, and they and departments increasingly used evaluations to inform expenditure management activities such as spending proposals (in particular program renewals) and spending reviews. At the same time, evaluations often did not meet central agencies' needs for information about program efficiency and economy.
4. Evaluation use under the 2009 policy was extensive, but use and impact could be improved by ensuring that the evaluations undertaken, and their timing, scope [1](#) and focus, [2](#) closely align with the needs of users.
5. The main use of evaluations was to support policy and program improvement.
6. The increased use of evaluations to support decision making was enabled by an observed government-wide culture shift toward valuing and using evaluations.
7. The factors that had the most evident positive influence on evaluation use in departments were policy elements related to governance and leadership of the evaluation function, whereas the factors that most

evidently hindered evaluation use were those related to resources and timelines.

8. Despite concerns about their capacity to meet all policy requirements, departments generally planned and expected to meet all requirements in the current five-year period.

Relevance and Impact of the Three Major Policy Requirements

Regarding the relevance and impact of the three major policy requirements (comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues ³), the evaluation found the following:

9. Challenges in implementing comprehensive coverage stemmed from the combined demands of the three key policy requirements (comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues), along with the context of limited resources for conducting evaluations. The five-year frequency requirement appeared to be central to the implementation challenges in most departments.
10. Stakeholders at all levels recognized the benefits of comprehensive coverage for encompassing the needs of all evaluation users and for serving all purposes targeted by the policy. Nevertheless, there were clear situations where individual evaluations had low utility.
11. The five-year frequency for evaluations demonstrated benefits and drawbacks that varied according to the nature of programs and the needs of users. To optimize evaluation utility for a given program, a longer, shorter or adjustable frequency may be required.
12. In combination with the comprehensive coverage requirement, the five-year frequency limited the flexibility of evaluation units to respond to emerging or higher-priority information needs.
13. In general, the five core issues covered the appropriate range of issues and provided a consistent framework that allowed for comparability and analysis of evaluations within and across departments, as well as across time. However, the perceived pertinence of some core issues varied by evaluation and by type of evaluation user.
14. Longstanding inadequacies in the availability and quality of program performance measurement data and incompatibly structured financial data continued to limit evaluators in providing assessments of program effectiveness, efficiency (including cost-effectiveness) and economy. Central agencies and senior managers desired, in particular, more and better information on program efficiency and economy.

Approaches Used to Measure Policy Performance

Regarding the approaches used to measure policy performance, the evaluation found the following:

15. Mechanisms for measuring policy performance tracked the obvious uses of evaluations—those that were direct and more immediate—but did not capture the range of indirect, long-term or more strategic uses, and may not have given a robust perspective on the usefulness of evaluations.

Other Findings

The evaluation also found the following:

16. The requirements of the *Policy on Evaluation* and those of other forms of oversight and review, such as internal audit, created some overlap and burden.

Conclusions

The 2009 *Policy on Evaluation* helped the government-wide evaluation function play a more prominent role in supporting the Expenditure Management System. The policy also supported uses such as program and policy

improvement, accountability and public reporting. Strong engagement from deputy heads and senior managers in the governance of the evaluation function promoted the utility of evaluations, and the evaluation needs of deputy heads, senior managers and central agencies were well served. In some cases, but not systematically across departments, evaluation functions produced horizontal analyses that contributed to useful cross-program learning, informing improvements both to the program evaluated and to other programs and to the organization as a whole. However, in assessing program effectiveness, efficiency and economy in evaluations, departmental functions were often limited by inadequacies in the availability and quality of performance measurement data and by incompatibly structured financial data.

The findings showed that while there was a general belief that all government spending should be evaluated periodically, there was also a widely held view that the potential for individual evaluations to be used should influence their conduct. Further, the policy requirements for evaluation timing and focus did not leave sufficient flexibility for departmental evaluation functions to fully reflect the needs of users in evaluation planning or to respond to emerging priorities. Evaluation needs were found to vary among different user groups (in particular, the needs of central agencies and departments were somewhat different). However, to fulfill coverage requirements within their resource constraints, departments sometimes chose evaluation strategies (for example, clustering programs for evaluation purposes) that were economical but that ultimately served a narrower range of users' needs. The lack of flexibility in the coverage and frequency requirements also made it challenging for departments to coordinate evaluation planning with other oversight functions in order to maximize the usefulness of evaluations and minimize program burden.

Recommendations

The evaluation recommends that when developing a renewed *Policy on Evaluation* for approval by the Treasury Board, the Treasury Board of Canada Secretariat should:

1. Reaffirm and build on the 2009 *Policy on Evaluation*'s requirements for the governance and leadership of departmental evaluation functions, which demonstrated positive influences on evaluation use in departments.
2. Add flexibility to the core requirements of the 2009 *Policy on Evaluation* and require departments to identify and consider the needs of the range of evaluation user groups when determining how to periodically evaluate organizational spending (including the scope of programming or spending examined in individual evaluations), the timing of individual evaluations, and the issues to examine in individual evaluations.
3. Work with stakeholders in departments and central agencies to establish criteria to guide departmental planning processes so that all organizational spending is considered for evaluation according to the core issues; that the needs of the range of key evaluation users, both within and outside the department, are understood and used to drive planning decisions; that the planned activities of other oversight functions are taken into account; and that the rationale for choices related to evaluation coverage and to the scope, timing and issues addressed in individual evaluations is transparent in departmental evaluation plans.
4. Engage the Secretariat's policy centres that guide departments in the collection and structuring of performance measurement data and financial management data in order to develop an integrated approach to better support departmental evaluation functions in assessing program effectiveness, efficiency and economy.
5. Promote practices, within the Secretariat and departments, for undertaking regular, systematic cross-cutting analyses on a broad range of completed evaluations and using these analyses to support organizational learning and strategic decision making across programs and organizations. In this regard,

the Treasury Board of Canada Secretariat should facilitate government-wide sharing of good practices for conducting and using cross-cutting analyses.

1.0 Introduction

1.1 Purpose of the Evaluation of the *Policy on Evaluation*

The 2009 [*Policy on Evaluation*](#) requires its own evaluation every five years.

The objectives of the evaluation were to:

- Assess the application and performance (effectiveness, efficiency and economy) of the policy and develop a baseline of results—in particular, those related to the use and utility of evaluation; and
- Identify opportunities to better support departments in meeting their evaluation needs through flexible application of the policy and the associated directive and standard.

This evaluation will inform the Treasury Board of Canada Secretariat in fulfilling its responsibilities to develop policy and to lead the government-wide evaluation function.

1.2 Background and Context

1.2.1 Evolution of the Federal *Policy on Evaluation* and the Context for Policy Renewal in 2009

The federal government has had central evaluation policies in place since 1977. Before the 2009 *Policy on Evaluation*, federal policies positioned the evaluation function to inform the management and improvement of programs, primarily from a program manager's perspective. In response to the increasing need for neutral, credible evidence on the value for money of government spending, the 2009 policy broadened the policy focus to include a more prominent role for the evaluation function in supporting the Expenditure Management System. Further, the policy situated the head of evaluation as a strategic advisor to the deputy head on the relevance and performance of departmental programs. Factors that led to refocusing the evaluation function included:

- The 2006 legislated requirement ([*Financial Administration Act, section 42.1*](#)) for all ongoing programs of grants and contributions to be reviewed for relevance and effectiveness every five years;
- The 2007 renewal of the Expenditure Management System, which was aligned with the Auditor General's recommendations of November 2006, ⁴ Budget 2006 commitments, and recommendations of the Standing Committee on Public Accounts ⁵ (adopted by the Standing Committee in February 2008 ⁶) on positioning evaluation to better support expenditure management decision making; and
- The advent of strategic and other spending reviews, which increased the demand for evaluations to provide information about program relevance and performance.

Coverage requirements existed in all previous federal evaluation policies and ranged from ensuring that all programs were evaluated periodically to considering, but not requiring, evaluation of all programs. The 2009 policy requires evaluations of all direct program spending. ⁷ Similarly, a frequency of evaluation was consistently specified in federal evaluation policies; however, this frequency varied from every three years to every six years. It is now every five years under the 2009 policy. Further, all federal evaluation policies included a set of issues to be addressed in evaluations. Since 1992 these issues have been consistent, requiring evaluations to examine the relevance, effectiveness and efficiency of programs. However, a notable

change in 2009 was that core evaluation issues were no longer discretionary; the 2009 policy makes it mandatory for evaluations to address five core issues in order to meet coverage requirements.

For more information on the evolution of evaluation in the federal government, see [Appendix A](#).

1.2.2 The International Context for Evaluation

Internationally, as governments undertook cost-cutting and cost-containment exercises in recent years, several countries expanded evaluation coverage and took steps to improve evaluation quality and to emphasize the use of evaluation in decision making. For example, the United Kingdom and the United States took steps to bolster the use of evaluation evidence in determining whether program spending is effective and provides value for money. The United Kingdom's guidance on evaluation for federal departments and agencies ⁸ indicates that with specific exceptions all policies, programs and projects should be comprehensively evaluated and that the risk of not evaluating is not knowing whether interventions are effective or delivering value for money. In the United States, evaluations are promoted as a means to “help the Administration determine how to spend taxpayer dollars effectively and efficiently—investing more in what works and less in what does not.” ⁹

1.2.3 Introduction of the 2009 Policy on Evaluation

The current federal *Policy on Evaluation* was introduced on April 1, 2009, replacing the 2001 *Evaluation Policy*. The objective of the 2009 policy is to create a comprehensive and reliable base of evaluation evidence that is used to support policy and program improvement, expenditure management, Cabinet decision making and public reporting. To meet this objective, the policy strengthened requirements for evaluation coverage; for the assessment of the value for money of programs; for the quality and timeliness of evaluations and the neutrality of the function; and for evaluation capacity in departments. In its [September 2010 report on Chapter 1, “Evaluating the Effectiveness of Programs,” of the Fall 2009 Report of the Auditor General of Canada](#), the Standing Committee on Public Accounts expressed support for the direction of the new policy by stating, “Effectiveness evaluations are very important for making good, informed decisions about program design and where to allocate resources. The Committee has long encouraged the development of effectiveness evaluation within the federal government and is pleased that the government has strengthened the requirements for evaluation.”

The 2009 policy and the associated [directive](#) and [standard](#) do the following:

- Establish evaluation as a function led by the deputy head, with a neutral departmental governance structure;
- Require comprehensive coverage of direct program spending ¹⁰ every five years;
- Articulate core issues of program relevance and performance that must be addressed in all evaluations (see [Appendix A, Table 2](#));
- Introduce requirements for program managers to develop and implement ongoing performance measurement strategies;
- Set competency requirements for departmental heads of evaluation;
- Set quality standards for individual evaluations; and
- Require that evaluation reports be made easily available to Canadians in a timely manner.

The 2009 *Policy on Evaluation* and *Directive on the Evaluation Function* introduced flexibilities to help departments achieve comprehensive coverage.

Because of the significant changes introduced by the policy, and on the advice of an advisory committee of deputy heads ¹¹ in 2008, a four-year phased implementation was adopted to give departments time to

build their capacity for achieving comprehensive evaluation coverage. During this transition period, departments could use a risk-based approach to choose which components of direct program spending to evaluate. The transition period did not apply to ongoing programs of grants and contributions, which had to be evaluated every five years in accordance with the 2006 legal requirement.

Following the transition period, which ended on March 31, 2013, all direct program spending became subject to evaluation, and by March 31, 2018, the requirement for comprehensive coverage will need to be met for the first time. As stipulated in [Annex A of the Directive on the Evaluation Function](#), departments could consider risk, program characteristics and other factors ¹² when choosing evaluation approaches and when calibrating the methods and the level of effort applied to each evaluation. For example, calibrating an evaluation to expend less effort could entail:

- Selecting fewer and more targeted evaluation questions to examine the core value-for-money issues, or to focus on known problem areas of the program;
- Choosing a streamlined evaluation approach and a design with a shortened timeline;
- Calibrating the methods used and the level of effort by leveraging existing data instead of collecting new data whenever possible; by using smaller sample sizes; by using lower-cost interviewing methods (such as online or telephone instead of in-person, or clusters of in-person interviews to limit travel costs); or by conducting fewer case studies.

Departments could also adjust the scope of evaluations by grouping programs rather than evaluating programs individually.

1.3 Overview of the Federal Evaluation Function

Under the 2009 *Policy on Evaluation*, evaluation serves various users, including deputy heads, central agencies, program managers, ministers, parliamentarians and Canadians. Evaluations support various uses, including policy and program improvement, expenditure management, Cabinet decision making and public reporting.

As examples of users and uses, evaluations may inform program managers about improvements to programs and proposals for program renewal or redesign (including Treasury Board submissions); support deputy heads in allocating resources across programs; support central agencies in playing their “challenge function” as they analyze and provide advice on Treasury Board submissions, Memoranda to Cabinet, and spending review proposals; and assist departments in reporting to parliamentarians and Canadians on program results.

Responsibilities for establishing and sustaining a strong federal evaluation function are shared. While the responsibility for conducting evaluations rests with individual departments and agencies, the Secretary of the Treasury Board plays a leadership role for the whole function, supported by the Centre of Excellence for Evaluation of the Treasury Board of Canada Secretariat. In leading the government-wide evaluation function, the Secretariat:

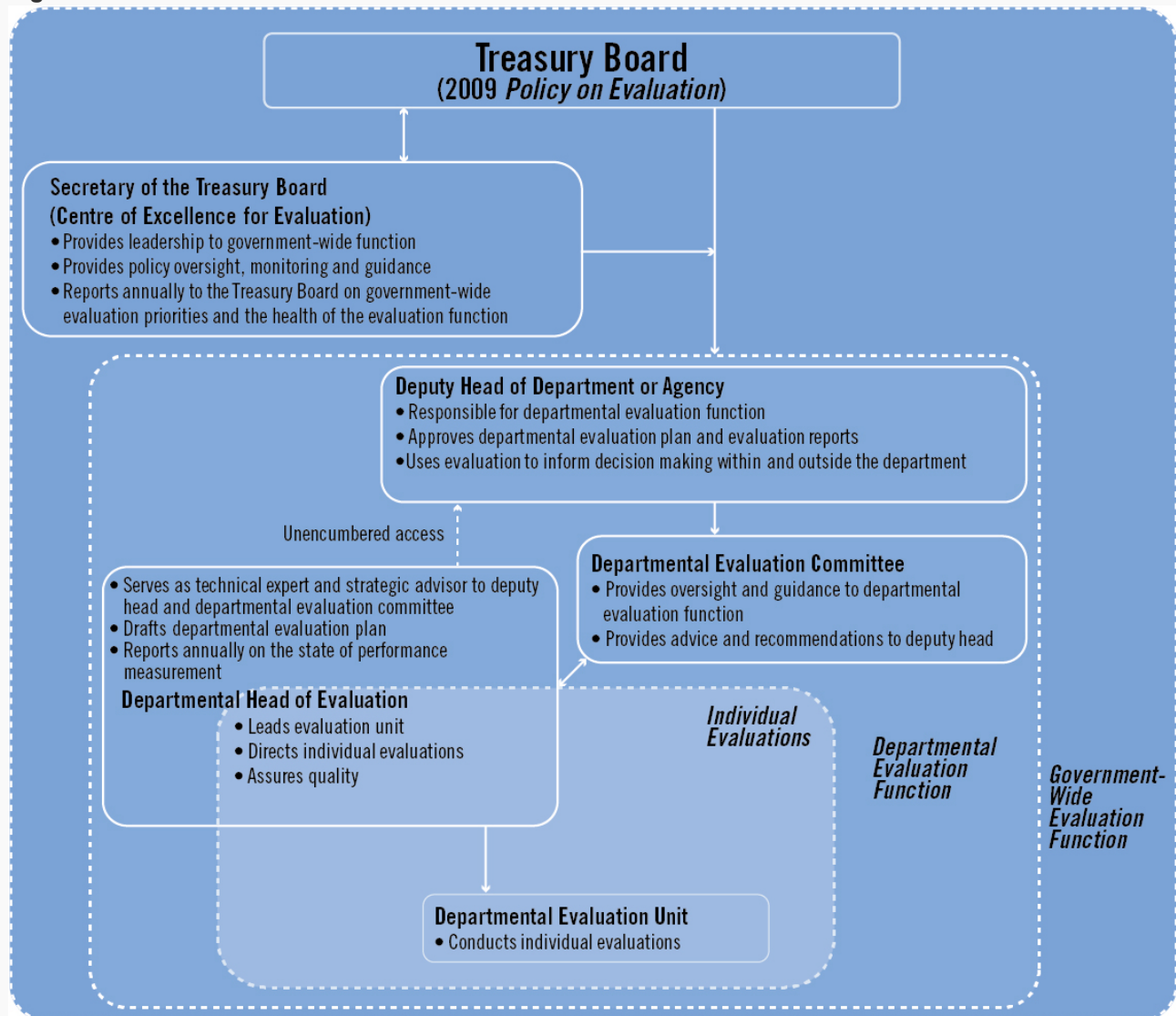
- Supports departments in implementing the 2009 *Policy on Evaluation*;
- Encourages the development and sharing of effective evaluation practices across departments;
- Supports capacity-building initiatives in the government-wide evaluation function;
- Monitors and reports annually to the Treasury Board on government-wide evaluation priorities and the health of the evaluation function; and
- Develops policy recommendations for the Treasury Board.

The *Policy on Evaluation* mandates key roles and structures for leading and governing departmental evaluation functions, as well as tools for planning their activities. These include the role of the head of

evaluation as the departmental lead for evaluation and strategic advisor to the deputy head; the role of the departmental evaluation committee in advising the deputy head and facilitating the use of evaluation; and the departmental evaluation plan as a tool for expressing plans and priorities and assisting the coordination of evaluation and performance measurement needs. In small departments and agencies, deputy heads lead the evaluation function. They are required to designate a head of evaluation, but they are not required to establish departmental evaluation committees or to develop departmental evaluation plans.

Figure 1 depicts the structure of the federal evaluation function and key roles and responsibilities, from the perspective of a large department or agency.

Figure 1. Structure of the Federal Evaluation Function



▼ **Figure 1 - Text version**

The figure shows the structure of the federal evaluation function in a large department or agency, including the governance structures and responsibilities of individuals and organizations, from three nested perspectives. The three perspectives, from narrowest to widest, are: individual evaluations, individual departmental evaluation functions and the government-wide evaluation function.

From the perspective of individual evaluations, departmental evaluation units are responsible for conducting evaluations. As the leader of the departmental evaluation unit, the departmental head of

evaluation directs individual evaluations and assures their quality.

From the perspective of the departmental evaluation function, the deputy head has overall responsibility for the departmental evaluation function, approves the departmental evaluation plan and individual evaluation reports, and uses evaluations to inform decision making within and outside the department or agency. The departmental evaluation committee provides oversight and guidance to the departmental evaluation function and provides advice and recommendations to the deputy head. The departmental head of evaluation serves as a technical expert and a strategic advisor to the deputy head and the departmental evaluation committee, drafts the departmental evaluation plan, and reports annually on the state of performance measurement. The departmental head of evaluation has unencumbered access to the deputy head on evaluation matters.

From the government-wide perspective, the Treasury Board sets government-wide policy for the federal evaluation function through the 2009 *Policy on Evaluation*. The policy establishes the role and responsibilities for the Secretary of the Treasury Board for providing leadership to the government-wide function. Supported by the Centre of Excellence for Evaluation, the Secretary provides policy oversight, monitoring and guidance, and reports annually to the Treasury Board on government-wide evaluation priorities and the health of the evaluation function.

1.4 Implementation of the *Policy on Evaluation*

After the introduction of the 2009 policy, the Treasury Board of Canada Secretariat continually monitored and reported on the policy's implementation. To identify issues, the Secretariat completed an Implementation Review in 2013 that examined the four-year policy transition period before full implementation of five-year comprehensive evaluation coverage.

Taken together, the Implementation Review and the Secretariat's Annual Reports on the Health of the Evaluation Function from 2010 to 2012 showed that departments had made solid progress during the policy's four-year transition period in establishing governance structures for the function (for example, departmental evaluation committees and heads of evaluation), building evaluation capacity, increasing evaluation coverage, planning for comprehensive coverage, and using evaluation to support decision making.

When introducing the 2009 policy, the Secretariat projected that departments would need to increase investment in the evaluation function to achieve and sustain comprehensive evaluation coverage every five years; however a period of government-wide spending reviews followed. Table 1 shows the number of evaluations and the resources allocated to them during the last two years of the 2001 policy and the four-year transition period of the 2009 policy, for large departments and agencies in the Government of Canada.

The Secretariat's monitoring showed that government-wide financial resources for the function were roughly stable until 2011–12 and then declined. However, the number of full-time equivalents dedicated to the function rose slightly compared with 2008–09 (the last year of the 2001 policy), apparently by reallocating budgets for professional services to salaries.

Table 1. Evaluation Functions of Large Departments and Agencies ^{*}, 2007–08 to 2012–13: Number of Evaluations, Full-Time Equivalents and Financial Resources ^{}**

	2001 Evaluation Policy		2009 Policy on Evaluation (transition period)			

Fiscal year	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13
Number of evaluations	121	134	164	136	146	123
Full-time equivalents	409	418	474	459	477	459
Financial resources (\$ millions)						
Salary	28.4	32.3	37.1	38.2	39.0	40.8
Operations and maintenance	17.9	4.4	5.0	4.3	4.6	3.8
Professional services	4.2	20.5	19.1	17.6	14.3	11.6
Other	6.7 [†]	3.7 ^{††}	5.8 ^{††}	0.3 ^{††}	2.2 ^{††}	Not applicable §
Total financial resources §§	57.3	60.9	66.9	60.2	60.2	56.2

Source: Capacity Assessment Surveys and Treasury Board of Canada Secretariat monitoring.

*
—

Includes organizations defined as large departments and agencies under the *Policy on Evaluation*, as determined each fiscal year. The list of large departments and agencies may vary from one year to the next.

**
—

Resource figures represent combined ongoing and time-limited resources.

†

For 2007–08, “other” includes other evaluation resources not managed by the head of evaluation as well as time limited resources for salary, operations and maintenance, and professional services.

††

For 2008–09 to 2011–12, “other” includes other evaluation resources not managed by the head of evaluation.

§

Starting in 2012-13, other resources were no longer monitored because they were not managed by the heads of evaluation.

§§

Figures may not add up to totals due to rounding.

Although the number of evaluations produced in the final year of the four-year transition period (123) was less than the number produced in the first year (164), evaluation reports produced in 2012–13 covered greater amounts of direct program spending than they had covered before 2009–10. In 2012–13, the average evaluation covered approximately \$78 million in direct program spending, compared with an average of \$44 million covered per evaluation in 2008–09. Thus evaluation information was available for a greater amount of direct program spending government-wide under the 2009 policy than under the 2001 policy.

The Implementation Review found that departments used one or more of the following strategies to expand evaluation coverage within their budgeted resources, including:

- Clustering programs for evaluation purposes;
- Calibrating the effort devoted to evaluation projects;
- Relying more on internal evaluators; and
- Minimizing non-evaluation activities.

For a summary of the findings from the Implementation Review, see [Appendix B](#).

1.5 Context for Policy Renewal in 2014

The evaluation of the 2009 *Policy on Evaluation* was carried out during the same period as a separately conducted assessment of the [Policy on Management, Resources and Results Structures](#). Together, these exercises provided input to a broader policy dialogue that sought opportunities for improving both policies.

2.0 Evaluation Approach and Design

2.1 Approach and Design

The evaluation used a largely goal-based learning approach aimed at determining the degree to which policy objectives were met and why, as well as a contribution analysis (theory-driven) model to identify and test the assumptions and mechanisms of the policy. The approach was also a collaborative one, in that the evaluation team included external consultants as well as analysts from the Treasury Board of Canada Secretariat's Centre of Excellence for Evaluation, which is the unit responsible for developing and making policy recommendations to the Treasury Board. The external team members assessed the performance of the *Policy on Evaluation*, established a baseline of results, and assessed the approaches that departments and the Secretariat used to measure the policy's performance. The internal team members examined the application of policy requirements and explored opportunities for adding flexibility. The external and internal evaluation team members provided a challenge function for each other's work and assured the quality of evaluation products.

The evaluation used various research designs, including multiple case studies, interrupted time series, retrospective pretests ¹³ and descriptive elements.

2.2 Methodology

The evaluation used the following methods:

- Policy performance case studies of 10 departments and agencies to qualitatively analyze evaluation use, using a total of 28 evaluations conducted across these departments. Eighty six key informant interviews were conducted with heads and directors of evaluation, evaluation team members, managers of evaluated programs, departmental evaluation committee members and central agency officials. Case studies also involved document reviews;
- Policy application case studies of six types of programs or categories of spending, using 24 examples from departments and agencies, to qualitatively analyze the relevance of key policy requirements and identify opportunities for flexibility in the requirements for comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues. For the case studies, 39 consultations were conducted with departmental program managers and evaluation

professionals, and 8 consultations were conducted with central agency representatives. The six types of programs or spending categories were:

- Assessed contributions to international organizations; ¹⁴
- Endowment funding; ¹⁵
- Programs with a requirement for recipient-commissioned independent evaluations; ¹⁶
- Low-risk programs;
- Programs with a long horizon to results achievement;
- Other programs identified by departments as challenging for policy application;
- Consultations with 35 heads of evaluation, or their delegates, in small group settings;
- Online surveys of 115 program managers and 153 evaluation managers and evaluators;
- Descriptive and inferential statistical analyses on policy monitoring data previously collected by the Centre of Excellence for Evaluation (Capacity Assessment Survey and Management Accountability Framework Assessment Results);
- Process mapping to give an overview of how the evaluation function operates in departments, including processes for planning, conducting and using evaluations;
- A review of internal and external documents, including the Implementation Review of the *Policy on Evaluation*, and a summary of consultations held in 2014 with deputy heads and other key respondents related to the five-year assessment exercises of the *Policy on Evaluation* and the *Policy on Management, Resources and Results Structures*; and
- A review of literature on the evaluation policies and practices of other jurisdictions, including the United States, the United Kingdom, Australia, Switzerland, Japan, India, South Africa, Mexico and Spain, as well as the United Nations Evaluation Group, the Development Assistance Committee of the Organisation for Economic Co-operation and Development, and the World Bank Independent Evaluation Group.

For more information on the methods used for the evaluation, see [Appendix C](#).

2.3 Governance

The evaluation was governed by two advisory committees: one composed of heads of evaluation, and the other composed of central agency representatives. Each committee's work was governed by terms of reference. Each committee provided comments and feedback on the overall evaluation plan, including evaluation questions and case study categories, the evaluation work plan for the external evaluation team, the preliminary findings of both the policy performance and policy application case studies, the draft overall findings for the final evaluation report, and the final evaluation report.

For more information on the governance committees for this evaluation, see [Appendix C](#).

2.4 Evaluation Period and Questions

The evaluation of the 2009 *Policy on Evaluation* covered the period since the policy was introduced on April 1, 2009, to March 31, 2014.

The evaluation questions were the following:

1. Under what circumstances or conditions, if any, is it appropriate to not address all five core issues in an evaluation? What impacts would this have on the use and utility of evaluations for different users (including those in line departments and central agencies) and the objective of the policy?

2. Under what circumstances and conditions, if any, is the five-year requirement for evaluation not appropriate? What impacts, if any, would changes to the five-year requirement have on the use and utility of evaluations for different users (including those in line departments and in central agencies) and the objective of the policy?
3. Is the comprehensive coverage approach the most appropriate model for ensuring that evaluation supports policy and program improvement, expenditure (direct program spending) management, Cabinet decision making, and public reporting?
4. To what extent are the current approaches to measuring policy performance appropriate, valid and reliable?
5. What are the baseline results for measures of policy outcomes specific to the use of evaluations to support:
 - Policy and program development and improvement?
 - Expenditure (direct program spending) management?
 - Cabinet decision making?
 - Accountability and public reporting?
 - Meeting the needs of deputy heads and other users of evaluation?
6. Are evaluations leading to improved expenditure (direct program spending) management decision making, effectiveness, efficiency or savings for programs and policies?
7. To what extent can outcome achievement be maintained given current capacity and resources?
8. What are the major internal and external factors influencing the achievement (or non achievement) of intended outcomes?

2.5 Limitations

The Centre of Excellence for Evaluation was both the manager of the entity under evaluation (the *Policy on Evaluation*) and a part of the evaluation team. To mitigate any concerns about the centre's objectivity in conducting the evaluation, the advisory committees reviewed evaluation plans and draft deliverables; the external team played a challenge role related to the work of the internal team; a quality assurance process was established for technical and final reports, for which the external and internal evaluation teams were both responsible; and a contribution theory was used for the *Policy on Evaluation* (see [Appendix D](#)) to analyze potential alternative explanations for observed policy outcomes.

For the case studies, departments self-identified evaluation examples, leading to a possibility of selection and response bias in the information provided about the examples. To validate the self-reported information, a review and analysis of documents and supporting literature was conducted by the evaluation team. In addition, consultations were held with central agency representatives and follow-up consultations were held with departmental representatives from both the evaluation unit and program areas.

In most cases, central agency representatives were not able to comment on specific cases (program examples or case study categories), as staff turnover had occurred since the completion of evaluations. Whenever possible, evidence related to the specific cases was gathered; otherwise, general perceptions and observations were explored on the applicability and utility of the policy requirements and on alternative approaches. In some cases, departments had also experienced turnover or did not respond to requests for consultations.

A potential limitation for the performance case studies was that for recent evaluations conducted according to the requirements of the 2009 policy, not enough time would have passed for those evaluations to be fully used. To mitigate this limitation, evaluations that were completed before 2013 were included among the selected cases.

3.0 Findings

3.1 Performance of the Policy and Status of Policy Outcomes

3.1.1 Baseline Results for Policy Outcomes (evaluation questions 5 and 6)

1. Finding: In general, the evaluation needs of deputy heads and senior managers were well served under the 2009 *Policy on Evaluation*. Senior management was able to draw strategic insights to support higher-level decision making. At the same time, efforts to meet the policy's coverage requirements sometimes made evaluation units less able to respond to senior management's emerging needs.

Deputy heads who were consulted indicated that under the 2009 policy, their departments produced a good base of evaluations and had the capacity to use them. The performance case studies showed that evaluations met a range of deputy head needs, such as:

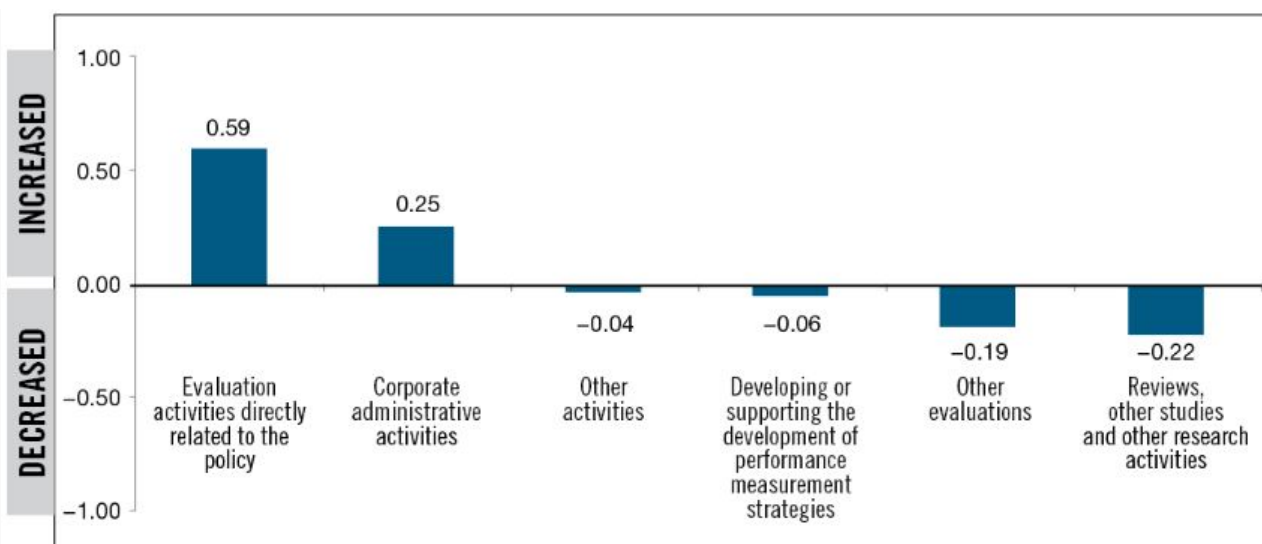
- Providing evidence of program effectiveness to support renewal decisions;
- Showing where program outcomes were not likely to be achieved; and
- Revealing related findings across a set of evaluations to support strategic decision making—for example, to identify an area of generalized concern.

The performance case studies showed that evaluations supported strategic decision making by delivering a more comprehensive perspective on the performance of departmental programming than that produced under the 2001 *Evaluation Policy*. The trend toward evaluating clusters of programs or larger entities, along with the convergence of all evaluations at departmental evaluation committees (or executive committees), enabled senior managers to recognize patterns across multiple evaluations and programs. Consultation evidence showed that some departments produced cross-cutting analyses from multiple evaluations of programs targeting common outcomes. In one case study, the insights drawn from across several evaluations led one deputy head to request a special review of a type of funding arrangement; in another case study, such insights influenced resource reallocation among a set of high-priority, horizontal activities. Senior executives on departmental evaluation committees also applied evaluation lessons from another branch to programs in their own branch.

Survey evidence showed that program managers felt that senior managers were well served under the 2009 policy. Three quarters of program managers surveyed (75%) reported that it was somewhat useful (38%) or very useful (37%) ¹⁷ for senior management (deputy ministers, associate deputy ministers and assistant deputy ministers) to have evaluations of their programs every five years, as required by the policy. Further, a majority of program managers (ranging from 68% to 87%) felt that each of the five required core issues was somewhat useful or very useful to senior management.

At the same time, the performance case studies showed that in some departments efforts to meet the policy's coverage requirements made evaluation units less able to respond to senior management's needs for special studies, reviews or specific evaluations on emerging issues. As shown in Figure 2, most evaluators surveyed reported that the proportion of time spent on evaluation activities directly related to the policy increased after the introduction of the 2009 policy, while the proportion of time spent on other evaluations, reviews, studies or research activities decreased.

Figure 2. Change in the Proportion of Time Evaluators Spent on Various Activities Since the Introduction of the 2009 *Policy on Evaluation* (N = 41 to 82)



Note: Mean values are based on responses on a three-point scale: -1.00 = decreased time spent, 0 = stayed the same, and 1.00 = increased time spent.

▼ Figure 2 - Text version

The figure shows the change in the proportion of time that evaluators, on average, have spent on various activities since the introduction of the 2009 *Policy on Evaluation*. The mean values for each of six activities are plotted on a vertical scale. The values from zero to one represent activities where an increased proportion of time was spent, and the values from zero to negative one represent activities where a decreased proportion of time was spent. The mean values are based on evaluators' survey responses, using a three-point scale where -1.00 means a decrease in the proportion of time spent, zero means the proportion of time spent stayed the same, and +1.00 means an increase in the proportion of time spent.

Two activities had mean values showing an increase, on average across evaluators, in the proportion of time spent on evaluation activities directly related to the policy (mean value of 0.59) and on corporate administrative activities (mean value of 0.25). Two activities had mean values that showed only a slight decrease, on average across evaluators, in the proportion of time spent on other activities (mean value of -0.04) and on developing or supporting the development of performance measurement strategies (mean value of -0.06). The remaining two activities had mean values that showed a decrease, on average across evaluators, in the proportion of time spent on other evaluations (mean value of -0.19) and on reviews, other studies, and other research activities (mean value of -0.22). The sample size (number of evaluators reporting) ranged from 41 to 82 depending on the activity.

2. Finding: The policy had an overall positive influence on meeting program managers' needs, and first-time evaluations of some programs were useful. However, program managers whose programs were evaluated as part of a broad program cluster or high-level Program Alignment Architecture entity sometimes found that their needs were not met as well as before 2009, when the program was evaluated on its own.

Program managers surveyed felt that evaluations were useful for a variety of purposes. In particular, 81% of program managers rated evaluations as somewhat useful (25%) or very useful (56%) for supporting program improvement, and 79% of program managers rated evaluations as somewhat useful (33%) or very useful (46%) for program and policy development. Performance case studies showed that some managers of

programs that were evaluated for the first time gained insights that led to improvements. Further, evidence from case studies suggested that these programs may never have been evaluated were it not for the policy's comprehensive coverage requirement.

At the same time, other evidence from the Implementation Review showed that program managers did not always find their programs reflected in the findings of evaluations whose scopes aligned with Program Alignment Architecture entities (a common scope for evaluations) ¹⁸ or encompassed clusters of programs. In these cases, the evaluations did not equip them with sufficiently detailed information to make program improvements. Performance case studies illustrated that some departments addressed this issue by designing these evaluations to produce findings and conclusions at multiple levels.

In terms of assisting program managers as they developed performance measurement strategies, the policy's influence was mixed. Based on the findings from the Implementation Review, the demands of the policy's comprehensive coverage and frequency requirements may have made some evaluation units too busy to support program managers in developing their strategies to the same extent that they once had. ¹⁹ However, some evaluation units emphasized their support to program managers in this regard, to ensure that performance measurement would support future evaluations. The survey of evaluators showed that following the introduction of the 2009 policy, a slightly larger proportion (38%) of evaluators decreased the time spent supporting the development of performance measurement strategies compared with the proportion (32%) that increased the time spent. The balance of evaluators indicated that the time spent stayed the same. Despite some evaluators spending less time supporting the development of performance measurement strategies, however, 90% of program managers surveyed in 2014 indicated that their programs had a performance measurement strategy in place. Among those programs with a performance measurement strategy in place, 93% of program managers had consulted their departmental evaluation function during its development.

3. Finding: Central agencies found that evaluations were increasingly available, and they and departments increasingly used evaluations to inform expenditure management activities such as spending proposals (in particular program renewals) and spending reviews. At the same time, evaluations often did not meet central agencies' needs for information on program efficiency and economy.

Central agency analysts generally viewed evaluations as a key source of program information and often consulted them first in their analysis of spending proposals. ²⁰ The performance case studies and stakeholder consultations showed that Secretariat analysts generally encouraged departmental use of evaluation findings in Treasury Board submissions, consistently required evaluation information for funding renewals in particular, and had recommended that departments not seek funding approval without a recent evaluation. Secretariat analysts reported that before the 2009 policy, evaluations were not always available to support submissions, but that today, if draft submissions do not provide evaluation information, they often seek such information from departments. In addition, when evaluation findings are negative, analysts seek departments' confirmation of corrective actions.

Several lines of evidence ²¹ indicated that evaluations were more widely used as a source of supporting information for Treasury Board submissions and, to a lesser extent, for Memoranda to Cabinet. Through the Capacity Assessment Survey, 96% of large organizations reported in 2012–13 that they used all or almost all relevant evaluations to inform Treasury Board submissions, and 78% reported that they used all or almost all relevant evaluations to inform Memoranda to Cabinet. These findings compare with those of the survey in 2008–09, prior to the 2009 policy, where 74% of large organizations reported that they almost always ²² considered evaluation results in Treasury Board submissions and 51% reported that they almost always considered them in Memoranda to Cabinet. Most large organizations established a formal process to include evaluation information in submissions (79%) and Memoranda to Cabinet (65%) in 2013–14. Evaluations were

commonly used to support renewals of existing spending—notably, for ongoing programs of grants and contributions. ²³ Central agency analysts typically used evaluation information to inform their advice to Treasury Board ministers, and some noted that they periodically received questions from Cabinet about evaluation results.

Based on performance and application case studies and on stakeholder consultations, the use and utility of evaluations, in particular at central agencies, was affected by how well evaluation timing aligned with the timing of spending decisions. Central agencies sometimes noted that evaluations arrived too late to meaningfully inform renewal decisions. For example, it was noted that key discussions on renewal are often held a year or more before a Treasury Board submission is developed. In those cases, an evaluation that is finished only in time to be appended to the submission can be seen as too late to support central agency analysts. It should also be noted, however, that within departments the draft evaluation reports are often available to program managers much earlier, allowing them to take advantage of the findings and knowledge generated, even if the report has not been fully approved.

Based on the Implementation Review and on case study consultations with central agency representatives, evaluation utility was also affected by how well the evaluation scope matched the unit of expenditure that was subject to a decision. When analyzing and advising on Memoranda to Cabinet or Treasury Board submissions, central agencies' information needs tended to be project-specific or program-specific—that is, specific to the unit of funding being renewed. When evaluations had a broad scope, such as a Program Alignment Architecture Program, they may not have provided sufficiently granular information. Case studies and stakeholder consultations showed that evaluations often did not meet central agencies' needs for information on program efficiency and economy—for example, because evaluators' analysis of program cost-effectiveness was limited by the incompatible structure of financial information. Central agencies also wanted better evidence on program alternatives in government-wide and cross-jurisdictional comparisons. A key risk associated with evaluations not meeting the needs of central agencies for this information is that their analysis and advice to ministers regarding departmental proposals may not be as well supported by neutral evidence as they could be.

Medium to high use ²⁴ of evaluations in spending reviews (for example, strategic reviews) was enabled by the increased availability and relevance of evaluations, ²⁵ and most departmental evaluation committee members and senior managers who were consulted, including deputy heads consulted in 2014, reported a high degree of evaluation utility for this purpose. Almost two thirds of evaluators surveyed reported positive impacts on the utility of evaluations for spending reviews because of the policy's comprehensive coverage requirement (63%) and core issues requirement (62%). Deputy heads consulted by the Treasury Board of Canada Secretariat in 2010 reported that strategic reviews raised the profile of the evaluation function by requiring departments to systematically address fundamental issues of program relevance. The majority of program managers surveyed (63%) reported that evaluations were somewhat useful (44%) or very useful (19%) for spending reviews. Most Secretariat program analysts ²⁶ reported that evaluations supported their analysis during spending reviews and noted that in many departments these reviews increased the demand for evaluations and that the attention paid by senior executives helped evaluation demonstrate its value.

4. Finding: Evaluation use under the 2009 policy was extensive, but use and impact could be improved by ensuring that the evaluations undertaken, and their timing, scope and focus, closely align with the needs of users.

Before the introduction of the 2009 policy, weaknesses in evaluation use had been documented. ²⁷ After 2009, the Secretariat's monitoring and reporting showed extensive evaluation use during the policy's transition period. In the 2012–13 Capacity Assessment Survey, large departments reported high

implementation rates of management responses and action plans; of the 901 management action plan items that were scheduled for completion in 2012–13, 53% were fully implemented by the end of the fiscal year and 21% were partially implemented. In addition, Management Accountability Framework assessment ratings documented extensive evaluation use; more than 96% of large departments were rated acceptable or strong for evaluation use in 2013–14, compared with 77% of large departments in 2007–08 and 78% of large departments in 2008–09.

When consulted in fall 2010, many deputy heads ²⁸ stated that evaluation was making a solid contribution to decision making, but a number of deputy heads felt that more could be achieved. When consulted in 2014, deputy heads acknowledged the usefulness of evaluations for program and policy improvement and development, strategic reviews and as a means of capturing corporate memory, while also noting that sometimes there were issues with the timing, focus and scale (level of intensity) of evaluations.

In case studies, evaluations were seen as most useful when they were timely, provided new information, and did not merely re-identify problems in program delivery that users already knew about. Evaluations were seen as less useful when they could not lead to organizational learning, when there was no decision to inform, or when no action could be taken. Central agencies as well as program managers, heads of evaluation, and evaluators noted situations where evaluations were less useful, including when their timing, scope, focus, report length and level of analytical rigour did not align with decision makers' needs or interests. Key risks associated with producing evaluations of low utility include spending evaluation resources inefficiently, rather than allocating the resources to evaluations that would be more useful and, more broadly, undermining the perceived value of the evaluation function as a whole.

5. Finding: The main use of evaluations was to support policy and program improvement.

Analysis by the Centre of Excellence for Evaluation showed that 75% of evaluation reports ²⁹ completed in 2010–11 included recommendations to improve program processes. Similarly, across the evaluations examined in the performance case studies, most recommendations pertained to program improvement, and all were actually used for this purpose. Performance case studies also showed that almost all of the evaluations examined were used for program improvement and that some evaluations resulted in improvements to a larger suite of programs than the one evaluated. ³⁰ However, performance case studies also showed that some evaluations were not used for improvement purposes when internal decisions left no opportunity for recommendations to be implemented—for example, when the program was eliminated or completely reorganized. The evidence showed that in the course of examining the relevance and performance of programs, evaluations sometimes contributed to operational efficiencies, but that these efficiencies were rarely in the form of direct cost savings.

Among a list of possible evaluation uses, program managers rated program and policy improvement as the one for which evaluations had been the most useful; 81% of program managers rated evaluations as somewhat useful (25%) or very useful (56%) for this purpose. Overall, both evaluators and program managers reported that the policy had a positive or neutral impact on the utility of evaluations for informing program and policy improvement; 56% of evaluators and 35% of program managers reported that evaluation utility increased, whereas only a small proportion (9% and 5% respectively) reported that utility had decreased. The balance (35% of evaluators and 60% of program managers) stated that utility had remained the same.

The evidence showed that the use of evaluation for accountability and public reporting increased, and program managers and evaluators reported that the policy had a positive impact on the utility of evaluations for these purposes. ³¹ The *2000 December Report of the Auditor General of Canada* noted that performance reports to Parliament made too little use of evaluation findings; by 2011, the annual Capacity

Assessment Survey showed that a high proportion of large organizations (89%) considered 80% or more of their evaluations when preparing their annual Departmental Performance Reports. In 2013–14, the annual Capacity Assessment Survey showed that 91% of large organizations had formal processes to ensure that evaluation inputs were considered in parliamentary reporting.

Performance case studies showed that organizations usually posted evaluation reports, including management responses and action plans, on their websites, although in some cases posting occurred long after the evaluation was completed. In the performance case studies, a small number of stakeholders suggested that part of the lag between functional completion of evaluation work and the approval and posting of reports was due to internal discussion when preparing reports for public posting, which led in some cases to less critical reporting.

6. Finding: The increased use of evaluations to support decision making was enabled by an observed government-wide culture shift toward valuing and using evaluations.

Key conditions had to be established for the policy to achieve its intended outcomes for evaluation use. ³² According to the theory of change developed for the *Policy on Evaluation* (see [Appendix D](#)), the policy was intended to drive a cultural shift in departments to increase the perceived value of, confidence in, and use of evaluation. This culture shift was evidenced by:

- The shift in stakeholder ³³ perceptions, from viewing evaluations as an oversight burden for programs, to perceiving their value and the skills available in the evaluation unit; ³⁴
- Increased departmental dialogue about evaluation since 2009, as reported by 71% of surveyed evaluators and 46% of surveyed program managers; ³⁵ and
- High rates of implementing recommendations, encouraged by the establishment of systems for tracking the implementation of evaluation recommendations, which 97% of all large departments reported having in place by 2013–14. ³⁶

Deputy heads' more general interest in evaluation likely contributed strongly to the observed outcomes. ³⁷ Management Accountability Framework assessment ratings of departmental evaluation functions ensured management attention and were partially responsible for bringing a higher profile to evaluation. An analysis of Management Accountability Framework assessment ratings from 2006–07 to 2011–12 showed an increasing trend in evaluation use, as well as coverage, governance and support, and quality of evaluation reports.

3.1.2 Factors Influencing Outcome Achievement (evaluation question 8)

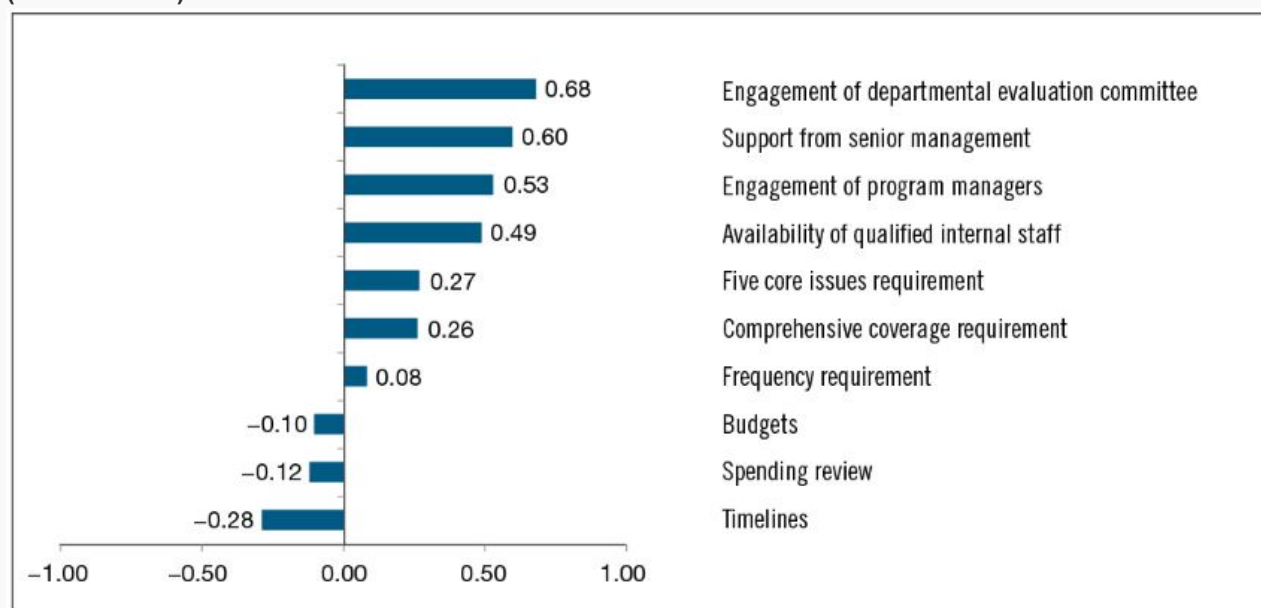
7. Finding: The factors that had the most evident positive influence on evaluation use in departments were policy elements related to governance and leadership of the evaluation function, whereas the factors that most evidently hindered evaluation use were those related to resources and timelines.

Across all lines of evidence, the engagement of senior leaders in departmental evaluation functions appeared to have the clearest positive influence on evaluation use. This influence was attributed, at least in part, to policy requirements related to governance and leadership (for example, the defined roles and responsibilities of deputy heads, departmental evaluation committees and heads of evaluation, and the head of evaluation's unencumbered access to the deputy head), and to a government-wide climate that emphasized results-based management and evidence-informed decision making. Increased senior management engagement led to greater implementation of action plans and enhanced the overall visibility of the evaluation function. The presence of deputy heads on most departmental evaluation committees ensured that evaluation findings were taken seriously, and scrutiny from a more senior executive level may have increased evaluation quality.

As shown in Figure 3, evaluators ³⁸ reported that the engagement of departmental evaluation committees, senior management and program managers, as well as the availability of qualified internal evaluation staff, had positive influences on achieving policy outcomes. These factors were more often reported by evaluators to have had a positive impact on use and utility than the policy requirements for comprehensive coverage of direct program spending, five-year frequency of evaluations, and examination of the five core issues.

Evaluators surveyed reported that the most evident negative influences on evaluation utility came from the timelines for evaluation projects, from the budgets for evaluation projects, and from spending reviews. Although the Treasury Board of Canada Secretariat did not track changes in the budgets of individual evaluation projects, government-wide financial resources for evaluation were 8% lower in 2012–13 than in 2008–09, before the policy's introduction, despite an initial increase in the first year of implementing the 2009 policy. The performance case studies provided further insights on the influence that spending reviews sometimes had as an external factor on evaluation utility: in some cases, evaluations could not be used for program improvement purposes because the program spending was significantly changed or discontinued. Program managers and senior managers also noted that the contribution of evaluations to decision making could be affected by the availability of other forms of oversight and review, such as spending reviews, especially when some of the input information was common.

Figure 3. Impact of Various Factors on Evaluation Use as Reported by Evaluators
(N = 98 to 141)



Note: Mean values are based on responses on a three-point scale: -1.00 = negative impact, 0 = no impact, and 1.00 = positive impact.

▼ Figure 3 - Text version

The figure shows the extent to which evaluators, on average, felt that various factors had positive or negative impacts on the use of evaluations. The mean values are based on evaluators' survey responses using a three-point scale, where -1.00 means the factor had a negative impact on evaluation use, zero means the factor had no impact on evaluation use, and +1.00 means the factor had a positive impact on evaluation use.

Seven factors had mean values indicating positive impacts on evaluation use: (1) engagement of departmental evaluation committee (mean value of 0.68), (2) support from senior management (mean value of 0.60), (3) engagement of program managers (mean value of 0.53), (4) availability of qualified internal staff (mean value of 0.49), (5) five core issues requirement (mean value of 0.27), (6)

comprehensive coverage requirement (mean value of 0.26), and (7) frequency requirement (mean value of 0.08). Three factors had mean values indicating negative impacts on evaluation use: (1) budgets (mean value of -0.10), (2) spending review (mean value of -0.12), and (3) timelines (mean value of -0.28). The sample size (number of evaluators reporting) ranged from 98 to 141 depending on the factor.

Management Accountability Framework assessments and ratings had a significant influence on policy implementation and results. On the positive side, they drew senior management's attention to evaluation and helped raise the profile of the function. On the negative side, they promoted risk-averse behaviour that may have limited departments' use of the policy's flexibilities. As noted in the Implementation Review, although flexibilities existed to calibrate evaluation effort when addressing core issues, they were not fully exploited owing to concerns that Management Accountability Framework assessment ratings would be adversely affected. This finding was corroborated by the performance case studies and stakeholder consultations, including consultations with deputy heads, who noted that while further policy flexibilities may be needed, existing flexibilities had not been fully exploited.

Another factor that influenced the policy's impact on the conduct and use of evaluations was the amount of grants and contributions spending administered by individual departments. In departments where the amount was large, the impact of the policy was small because of the pre-existing *Financial Administration Act* requirement (section 42.1) for comprehensive five-year coverage of this spending. Stakeholders noted that organizations with a large amount of grants and contributions spending had evaluation functions that were well established and producing useful evaluations before 2009.

3.1.3 Sustainability of Outcomes (evaluation question 7)

8. Finding: Despite concerns about their capacity to meet all policy requirements, departments generally planned and expected to meet all requirements in the current five-year period.

A comparison of Capacity Assessment Survey data collected before 2009 and in 2012–13 showed that on average, large organizations increased the human resources they devoted to their evaluation functions by 10% but decreased financial resources by 8%. As mentioned earlier, to expand evaluation coverage with these resources, departments used various strategies. Evaluators ³⁹ reported that the most effective strategies were calibrating evaluation scope and approach according to program risks, aligning an evaluation's scope with Program Alignment Architecture units, clustering related programs, and increasing the use of internal staff to conduct evaluations. ⁴⁰ The trend toward conducting evaluations with broad scopes, however, affected the utility of evaluations; for example, some program managers found that the information available to inform program improvements was less detailed.

When consulted for the Implementation Review, heads of evaluation highlighted resources as the main factor constraining them in meeting the coverage requirements, and they were concerned about their ability to meet the requirements in a meaningful manner with the available resources. Although heads of evaluation and other stakeholders ⁴¹ expressed concerns about the function's capacity to achieve and maintain comprehensive coverage over five years, in most cases it appeared that departments could manage their capacity to meet the requirements. Three quarters of evaluators ⁴² (74%) reported that the utility of the evaluation function could be maintained with current resources, and in a subsequent question, more than one third (36%) stated that utility could be increased.

Despite the potential for achieving full evaluation coverage within current capacity, several lines of evidence ⁴³ showed that greater flexibility is needed in applying policy requirements related to coverage, timing, scope and focus for evaluations to be more responsive to the information needs of various users. Flexibility is further discussed in subsequent sections of this report.

3.2 Application of the Three Major Policy Requirements

9. Finding: Challenges in implementing comprehensive coverage stemmed from the combined demands of the three key policy requirements (comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues), along with the context of limited resources for conducting evaluations. The five-year frequency requirement appeared to be central to the implementation challenges in most departments.

Although the relevance and impact of the three major policy requirements are discussed separately in the subsections below, this evaluation found that there was a clear interaction among the requirements. For example, challenges associated with comprehensive coverage were often related to the five-year time frame for completing comprehensive coverage or to the requirement for addressing all five core issues, rather than to the comprehensive coverage requirement on its own. Arguably, key challenges associated with the comprehensive coverage and the core issues requirements could be attributed in large measure to the five-year frequency requirement. Many stakeholders supported the periodic evaluation of all programs, but not the inflexibility of a five-year frequency when it did not meet their information needs. Others supported the principle of addressing core issues but questioned the need to address all of them every five years.

Most lines of evidence, ⁴⁴ including stakeholder consultations, documented doubts about departments' capacity to achieve comprehensive coverage over five years. Some stakeholders felt the requirements were too demanding given current resources; others identified resources as the key constraint to meeting coverage requirements and producing meaningful evaluations. In this context, the existing trend toward evaluating larger program entities ⁴⁵ was reinforced by the comprehensive coverage and five-year frequency requirements, which led departments to increasingly opt for evaluating programs in clusters or as Program Alignment Architecture units. Deputy heads consulted in 2014 stated that the comprehensive coverage requirement encouraged departments to evaluate larger units of programming, and case study evidence showed that departments commonly used this strategy to expand evaluation coverage. Further, a sample of departmental evaluation plans analyzed by the Centre of Excellence for Evaluation in 2011 showed that the scope of two thirds of evaluations aligned with Program Alignment Architecture Programs or Sub-Programs.

Case studies showed that to meet the coverage requirements, departments sometimes diverted evaluation resources from higher-priority work or emerging needs to evaluations of low-risk, small or unimportant programs. Deputy heads consulted in 2014 indicated that requiring comprehensive coverage to be achieved over a five-year period limited the flexibility of departments to target evaluations on new or emerging priorities. The policy's five-year frequency requirement, coupled with the *Financial Administration Act's* requirement for five-year coverage of all ongoing programs of grants and contributions, meant that in some organizations and in some years, the timing for completing evaluations had been inflexible for a high proportion of them. For example, one head of evaluation suggested that up to 80% of the unit's evaluation plan was fixed as a result of the requirements of the policy and the *Financial Administration Act*. It was noted in the consultations that little differentiation was made between the five-year requirement of the *Financial Administration Act* (section 42.1) that pertained specifically to ongoing programs of grants and contributions, and the coverage requirements of the *Policy on Evaluation*. This finding suggests that deputy heads' observations on the challenges and inflexibilities of the policy's five-year comprehensive coverage requirement may also apply to the legal requirement. ⁴⁶

A further combined challenge of the five-year frequency and comprehensive coverage requirements was that some evaluations had to be conducted when a program was immature or when its performance measurement data were insufficient, which made these evaluations less useful and more difficult to conduct.

Consultations and other evidence showed that despite the challenges of the coverage requirements, departments did not try to avoid conducting evaluations. However, they perceived a need for the Treasury Board of Canada Secretariat to allow departments to apply the three major policy requirements more flexibly, to ensure the value, utility, efficiency and cost-effectiveness of evaluation. A prevalent view among stakeholders was that for core issues, frequency and coverage, evaluations under the 2009 policy were intended to satisfy central agencies' information needs as much as or more than the needs of senior management in departments.

3.2.1 Comprehensive Coverage (evaluation question 3)

10. Finding: Stakeholders at all levels recognized the benefits of comprehensive coverage for encompassing the needs of all evaluation users and for serving all purposes targeted by the policy. Nevertheless, there were clear situations where individual evaluations had low utility.

A literature review showed that six ⁴⁷ of nine countries, as well as the Development Assistance Committee of the Organisation for Economic Co-operation and Development, recommended comprehensive evaluation coverage. Alternative approaches used in other jurisdictions involved targeting evaluation coverage by considering a variety of factors such as decision-making needs, priorities, program maturity, program type, self-assessment results and important government-wide themes.

Deputy heads who were consulted in 2014 held mixed opinions about the appropriateness of the comprehensive coverage requirement; many were in favour, some emphatically, whereas a smaller number favoured a more risk-based model. However, those who favoured the comprehensive coverage model often stated that the five-year period for achieving comprehensive coverage posed challenges. The reasons that deputy heads supported comprehensive coverage included the following:

- It makes sense to evaluate all programming.
- It ensures disciplined oversight, ensures accountability and keeps issues from being “swept under the carpet.”
- It leads to evaluation scopes that are often at a higher level and that support decision making on important “units of account.”

Central agency respondents generally favoured comprehensive coverage because it demonstrates good governance to scrutinize government's use of all taxpayers' contributions. Case studies showed that without the comprehensive coverage requirement some low-priority or low-risk programming would have been excluded from evaluation. However, central agency respondents felt it appropriate to periodically evaluate both low-risk and long-horizon programs, as these evaluations could be important in formulating advice to Treasury Board ministers on program renewals and for ensuring public accountability. The performance case studies demonstrated that there was value in evaluating some low-risk programs. At the same time, central agency respondents also recognized that evaluating certain programs was impractical. When central agencies had concerns about the comprehensive coverage requirement, the concerns often related to conducting evaluations of low utility—for example, evaluations that could not lead to actionable recommendations.

Heads of evaluation who were consulted generally agreed on several benefits that they observed from the comprehensive coverage requirement since 2009. In particular, they reported that it made evaluations available to inform decision making (notably, for programs where no past evaluations existed) and to inform

processes such as departmental performance reporting and adjustments to Program Alignment Architectures. They also reported that comprehensive coverage increased the profile of the function, validating and empowering the function within departments, while increasing its workload and sometimes its resources. ⁴⁸

The views of other stakeholder groups, notably program managers and evaluators, ⁴⁹ were more divided. Among those stakeholders that supported comprehensive coverage, there was a general view that in principle, all spending should be evaluated periodically. These stakeholders reported that benefits from the comprehensive coverage requirement included insights on programs never-before or long-ago evaluated, and a strategic view of performance that cut across related departmental programs—for example, to identify redundancies and synergies. Evaluators, in particular, generally reported that comprehensive coverage had increased the utility of evaluations for all major uses targeted by the policy. ⁵⁰ In addition, case studies showed that the requirement had a profound effect in some departments, especially those with little grants and contributions spending, because many evaluations were conducted on entities that had never been evaluated before 2009. These evaluations sometimes produced valuable findings that led to program improvements. In some cases, however, stakeholders indicated that if an evaluation had not been required, the organization would likely have conducted a different type of study to address its needs.

Across all stakeholder groups, those that did not support comprehensive coverage generally questioned using resources to evaluate programs where there was little perceived need for the information (for example, for low-risk programs, or where other sources of information existed), where the evaluation might have no utility or where recommendations would be non-actionable. For example, the case studies showed that some stakeholders questioned the merits of applying the comprehensive coverage requirement to assessed contributions because evaluations would have no impact on what Canada is required to spend on these programs. However, case studies showed that existing evaluations of assessed contributions, which focused on the effectiveness of Canada's membership effort (for example, in deriving benefits for Canada or in influencing organizational policies) or on the coordination between the various departments and agencies engaged with the international organization) had demonstrated value. Assessed contributions are also subject to the evaluation requirements of the *Financial Administration Act* (section 42.1) and the *Policy on Transfer Payments*.

A small number of deputy heads ⁵¹ and heads of evaluation and a minority of program managers and evaluators suggested returning to the former risk-based model for evaluation planning—that is, using risk considerations to decide whether to evaluate programs. Several caveats accompanied this suggestion, including that risk-based approaches are not a panacea; that clear guidance from the centre would be needed to ensure consistency in the assessment of risk; that materiality alone should not be used to define risk; and that at least one full round of comprehensive coverage could be needed to provide assurance that program risk levels are accurately assessed.

A number of alternative approaches to coverage were suggested in case studies and consultations, including:

- Comprehensive coverage over a longer time frame (longer frequency);
- Risk-based coverage, with calibration of individual evaluations;
- Targeted evaluations (evaluations that have narrower scopes or that address specific issues or themes);
- Evaluations focused on departmental priorities and interests; and
- Evaluations where there is the most potential for information gain.

In contrast, some central agency respondents and heads of evaluation recommended that the policy be expanded to require the evaluation of types of program spending that it currently does not evaluate. Specific program types mentioned were sunseting or time-limited programs, internal services and statutory programs

(beyond the administrative aspects alone). Central agencies in particular noted that sunseting and time-limited programs are sometimes renewed and that evaluations are helpful for informing the renewal process.

3.2.2 The Five-year Frequency for Evaluations (evaluation question 2)

11. Finding: The five-year frequency for evaluations demonstrated benefits and drawbacks that varied according to the nature of programs and the needs of users. To optimize evaluation utility for a given program, a longer, shorter or adjustable frequency may be required.

12. Finding: In combination with the comprehensive coverage requirement, the five-year frequency limited the flexibility of evaluation units to respond to emerging or higher priority information needs.

Within and across stakeholder groups, there were diverse perspectives on the appropriateness and utility of the policy's five-year frequency requirement. Although the five-year frequency for evaluations was perceived to be mechanistic and insensitive to management needs and preferences, all stakeholder groups noted both positive and negative aspects.

Compared with other stakeholder groups, central agencies more strongly supported the five-year frequency requirement. This support may be explained by the nature of central agency work, which often involves advising the Treasury Board on funding renewals that usually occur every five years and that can be informed by evaluations.

Heads of evaluation viewed the positives and negatives associated with the five-year frequency requirement with equal emphasis, whereas program managers [52](#) and evaluators [53](#) were more likely to favour alternative frequencies. For example, 63% of evaluators surveyed thought that a different frequency than "every five years" should be used for evaluating programs, excluding ongoing programs of grants and contributions. It should also be noted that the experiences of departments in implementing the 2009 policy may have varied widely owing to the nature of their programs, their evaluation history and the experiences of their deputy heads and heads of evaluation. [54](#)

Positive aspects of the five-year frequency requirement expressed by a majority of stakeholder groups included the following:

Availability and broad support to decision making:

The requirement created and ensured a constant foundation of information about all direct program spending to support all areas of decision making—for example, expenditure management needs, spending reviews and accountability.

Recent information:

The requirement ensured that evaluation information would be no older than five years. Central agencies [55](#) and senior executives said that they discounted evaluations older than five years and considered that those conducted in the past two or three years were the most useful—for example, for informing Treasury Board submissions.

Overall strengthening of the evaluation function:

The requirement supported a culture of evaluation, performance measurement and learning; increased the profile of evaluation functions; enabled evaluation planning; and increased engagement of program managers, including in performance measurement.

Negative aspects of the five-year frequency requirement that were identified by most stakeholder groups included the following:

Evaluation timing not always aligned with needs:

The requirement led to mechanistic timing that did not always produce an evaluation at the time it would be useful—for example, when program context was stable; when there was no specific decision to inform; when recent information was available from other sources or review activities; when recent significant program changes made evaluation premature; or when an earlier timing would provide useful formative information on a new program.

Strained evaluation capacity and reduced responsiveness:

The requirement lessened the responsiveness of some evaluation functions to new and emerging information needs, or limited their capacity to focus on the real problems, higher-value evaluation projects or strategic activities, as resources were fully committed to delivering on departmental evaluation plans. In some cases, this requirement led to prioritizing evaluations of low risk or low interest to deputy heads. It was also noted that the length of time to complete evaluations affects how easily evaluations can fit into a five-year cycle. [56](#)

Effects of broad evaluation scopes:

As described earlier, the requirement resulted in strategies to evaluate programs in clusters or as larger entities. This approach reduced the utility of evaluations for some users but resulted in increased utility for others.

Poor use of resources:

As reported by the Auditor General in spring 2013, [57](#) the requirement limited the ability of departments "...to put their evaluation resources to the best use." The majority of stakeholder groups expressed similar concerns.

Lower perceived value of the evaluation function:

Some reported a combined negative effect of the above (timing not aligned with needs, low responsiveness and poor use of resources) on the overall perceived value of the evaluation function.

In addition, heads of evaluation and evaluators reported a negative aspect of the five-year requirement to be the burden on programs from undergoing repeated evaluations. Despite this burden, three quarters (77%) of program managers surveyed whose programs had been evaluated since 2009 felt that it was somewhat useful (41%) or very useful (36%) to have an evaluation of their program every five years, to support their decision making and program management needs.

Although a regular, cyclical evaluation approach was often preferred over an ad hoc approach, the dominant view was that the five-year frequency was not appropriate in some situations. Many favoured the notion of having greater flexibility in evaluation frequency so that evaluations could be timed to ensure maximum utility, including delaying one evaluation in order to facilitate another. Where the five-year frequency was inappropriate, stakeholders proposed many alternative triggers and considerations for evaluation timing. Based on evidence from multiple sources, [58](#) alternative approaches fell into three broad categories:

- Adopting a longer evaluation frequency, such as seven, eight or even 10 years;
- Maintaining a five-year frequency while allowing departments to target evaluations on specific programs areas or smaller program components, rather than on full programs; or
- Determining evaluation timing based on needs.

Internationally, evaluation policies most commonly link evaluation timing to decision making and reporting processes or requirements. This evaluation found that the increased alignment of evaluation timing with

Treasury Board submissions contributed to the use and utility of evaluations.

Across all stakeholder groups, a proportion of stakeholders indicated that the importance of the program and the level of risk should be the main factors for determining evaluation frequency and that the life cycle or maturity of the program should also be considered. To help align evaluation timing with needs, several factors were identified across most stakeholder groups, including:

- Risk;
- Program factors—for example, materiality or importance, and type of funding;
- Program maturity or program life cycle;
- Decision-making needs;
- Known program issues;
- Planned restructuring; and
- Availability of performance measurement data or information from other sources.

Based on case study evidence, departments favoured the idea of choosing the appropriate evaluation frequency for their programs themselves, as this would help them meet ongoing and emerging needs (including their own, and those of central agencies); coordinate evaluation timing with other oversight activities (for example, audits and reviews); and minimize program burden. Some evidence suggested that central agencies (notably, the Treasury Board of Canada Secretariat) could play a role in informing departmental evaluation planning decisions on frequency or timing to ensure that expenditure management needs are met; however, there were no suggestions on how to put this into practice.

3.2.3 Core issues (evaluation question 1)

13. Finding: In general, the five core issues covered the appropriate range of issues and provided a consistent framework that allowed for comparability and analysis of evaluations within and across departments, as well as across time. However, the perceived pertinence of some core issues varied by evaluation and by type of evaluation user.

Specifying core issues is consistent with practices in the other jurisdictions examined. Further, the core issues identified in the 2009 *Directive on the Evaluation Function* are largely consistent with the categories of issues addressed in other jurisdictions (relevance, effectiveness and efficiency), although the breakdown of issues into components (for example, the breakdown of relevance into program need, government priority and government role) is less common.

Each of the policy's five core issues has one or more interested users within departments and central agencies and at the political and public levels. In terms of evaluation utility, central agencies generally benefited more than departments from the consistent application of all core issues. ⁵⁹ Documents from the Treasury Board of Canada Secretariat indicated that the five core issues were intended to focus evaluations on examining program value for money. In the analysis of case studies, evaluators, program managers and senior management (including departmental evaluation committee members) perceived the core issues to be appropriate and useful. Program managers, in particular, perceived the core issues to be the “right ones.” Central agencies generally favoured consistent application of all five core issues across evaluations because they aligned with the types of questions ministers ask and the questions central agencies ask in playing their challenge function role. Moreover, the consistent framework of evaluation issues could support central agencies' emerging uses for performance information—for example, horizontal analytics or syntheses of evaluation information. No issue was consistently identified as missing from the set of core issues.

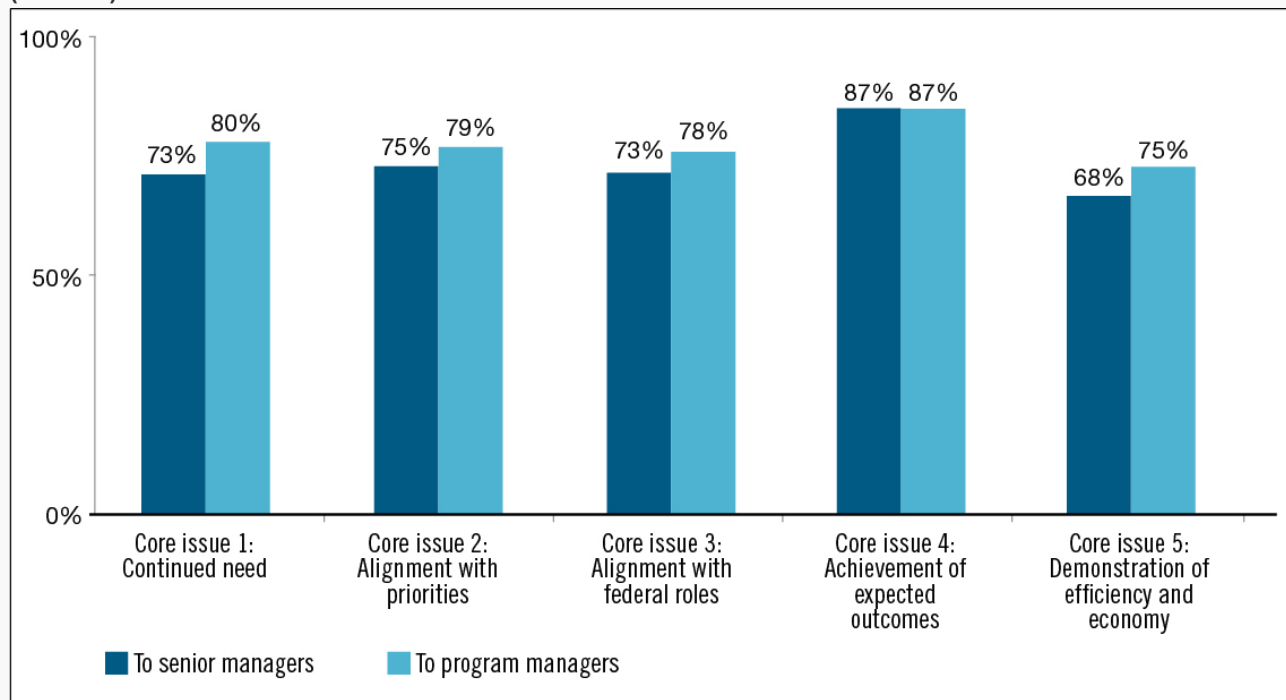
The core issues requirement had both positive and negative impacts on the utility of evaluations. ⁶⁰

The following are examples of the positive impacts:

- It helped support program improvement.
- It provided a consistent framework to focus evaluations and allow program comparisons—for example, systematic issues, overlaps and opportunities for synergy. This consistency supported the work of Secretariat analysts and prevented information gaps. [61](#)
- It focused evaluations on providing the information needed to support expenditure management (for example, supporting Treasury Board submissions and Memoranda to Cabinet), spending and program reviews, needs of decision makers and other evaluation users, and good management.
- It helped ensure that programs were not avoiding suspected performance issues.
- It highlighted problems with performance measurement data.

As shown in Figure 4, program managers surveyed generally viewed all core issues as somewhat useful or very useful. [62](#)

Figure 4. Percentage of Program Managers That Rated Each Core Issue as Somewhat Useful or Very Useful In Supporting Their Decision-Making Needs and Those of Senior Managers (N = 115)



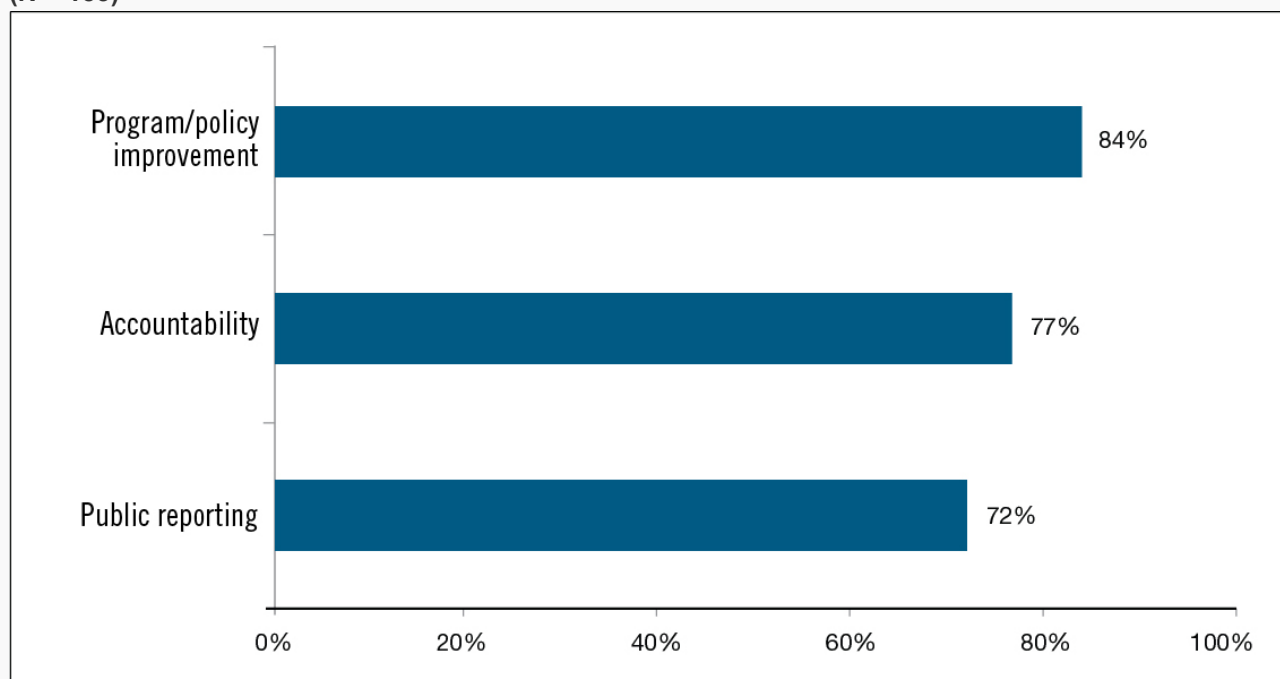
▼ Figure 4 - Text version

The figure shows the percentage of the 115 program managers surveyed that rated each of the five core issues as somewhat useful or very useful in supporting the decision-making needs of senior managers and the proportion that rated each core issue as somewhat useful or very useful in supporting their own decision-making needs. For core issue 1, “continued need,” 73% of program managers rated the issue as somewhat useful or very useful in supporting the decision-making needs of senior managers and 80% rated the issue as somewhat useful or very useful in supporting their own decision-making needs. For core issue 2, “alignment with priorities,” 75% of program managers rated the issue as somewhat useful or very useful in supporting the decision-making needs of senior managers and 79% rated the issue as somewhat useful or very useful in supporting their own decision-making needs. For core issue 3, “alignment with federal roles,” 73% of program managers rated the issue as somewhat useful or very useful in supporting the decision-making needs of senior managers and 78% rated the issue as

somewhat useful or very useful in supporting their own decision-making needs. For core issue 4, “achievement of expected outcomes,” 87% of program managers rated the issue as somewhat useful or very useful in supporting the decision-making needs of senior managers and 87% rated the issue as somewhat useful or very useful in supporting their own decision-making needs. For core issue 5, “demonstration of efficiency and economy,” 68% of program managers rated the issue as somewhat useful or very useful in supporting the decision-making needs of senior managers and 75% rated the issue as somewhat useful or very useful in supporting their own decision-making needs.

As shown in Figure 5, the majority of evaluators surveyed reported that the core issues requirement had the most positive impacts on the utility of evaluations for informing program improvement, and for accountability and public reporting.

Figure 5. Percentage of Evaluators That Reported Positive Impacts From the Core Issues Requirement on the Utility of Evaluations for Various Uses
(N = 153)



▼ Figure 5 - Text version

The figure shows the percentage of evaluators surveyed that reported that the core issues requirement had a positive impact on the utility of evaluations for each of three purposes. Of the 153 evaluators surveyed, 84% reported a positive impact on the utility of evaluations for program and policy improvement; 77% reported a positive impact on the utility of evaluations for accountability purposes; and 72% reported a positive impact on the utility of evaluations for public reporting.

Negative impacts associated with the core issues ⁶³ related to the limited flexibility that departments reported for designing evaluations to reflect user needs and priorities (including those of deputy heads ⁶⁴ and program managers) and program characteristics and context. ⁶⁵

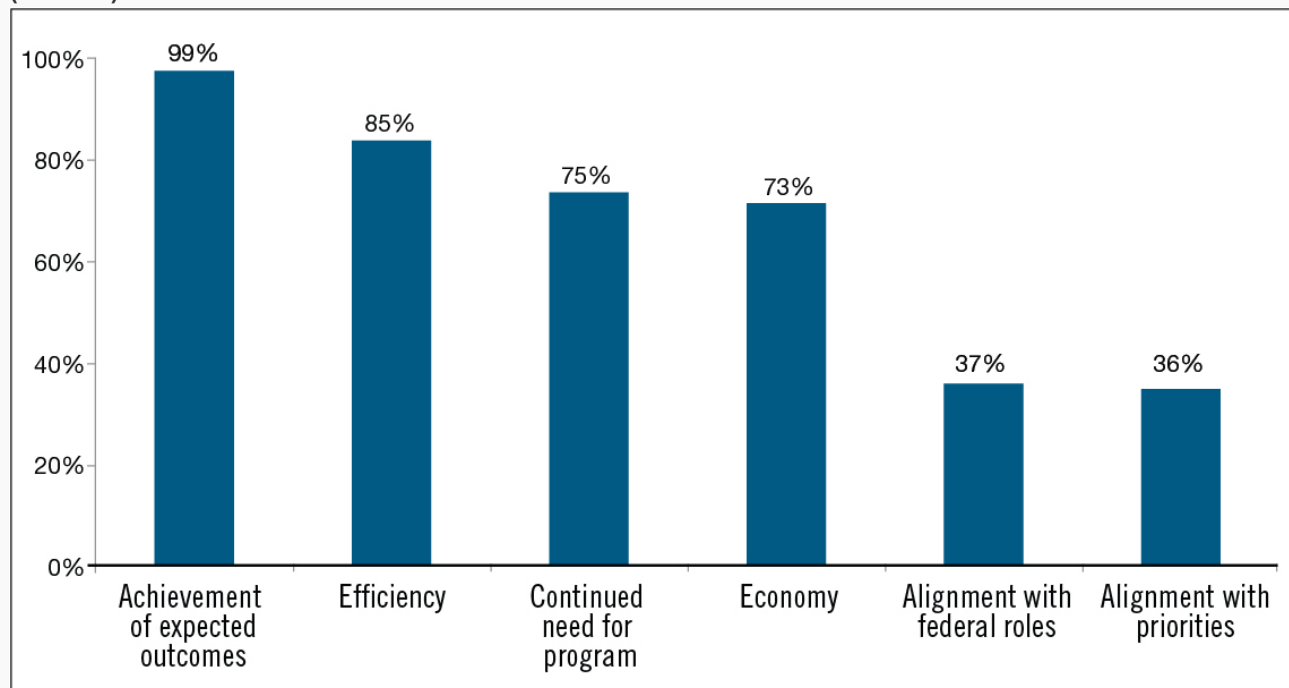
The general support for addressing core issues was accompanied by observations that periodically and in specific circumstances, certain core issues might not be applicable in a given evaluation. Most stakeholder groups agreed that the following three core issues were important or essential and should be addressed in all evaluations:

- Demonstration of efficiency and economy, which examines a program's use of resources in achieving outcomes and was reported by some stakeholders as being of highest importance for parliamentarians;
- Achievement of expected outcomes, which examines program effectiveness and is of great interest to program managers, senior managers and deputy heads; and
- Continued need for the program, which was reported to provide useful information for both program managers and senior managers.

The first two are core performance issues and were of greatest concern to deputy heads, program managers and other users of evaluation, [66](#) and the third is a core relevance issue.

As well, as shown in Figure 6, most evaluators felt that these same core issues should be addressed in all evaluations (note that the survey divided the “demonstration of efficiency and economy” core issue into two parts, to measure evaluators’ perceptions of each).

Figure 6. Percentage of Support by Evaluators for Inclusion of Each Core Issue Concept in All Evaluations
(N = 153)



▼ Figure 6 - Text version

For each of the five core issues, the figure shows the percentage of evaluators surveyed that felt the issue should be addressed in all evaluations. Of note, the survey divided the “demonstration of efficiency and economy” core issue into two parts. From the highest percentage to the lowest percentage, 99% of the 153 evaluators felt that the “achievement of expected outcomes” core issue should be addressed in all evaluations; 85% of evaluators felt that “efficiency” (part of the “demonstration of efficiency and economy” core issue) should be addressed in all evaluations; 75% of evaluators felt that the “continued need for program” core issue should be addressed in all evaluations; 73% of evaluators felt that

“economy” (part of the “demonstration of efficiency and economy” core issue) should be addressed in all evaluations; 37% of evaluators felt that the “alignment with federal roles” core issue should be addressed in all evaluations; and 36% of evaluators felt that the “alignment with priorities” core issue should be addressed in all evaluations.

Across the stakeholder groups, some stakeholders stated that two of the three relevance issues—alignment with federal roles and responsibilities, and alignment with government priorities—applied only to a subset of programs or only under certain conditions. ⁶⁷ Other stakeholders stated that all three core relevance issues were essential or the most important, especially for informing senior management. Senior management and central agency respondents indicated that the relevance issues helped address questions in the expenditure management review framework and, in particular, the question of affordability.

When alignment with federal roles and responsibilities and alignment with government priorities were perceived to be inapplicable, ⁶⁸ stakeholders felt that addressing them would be an inefficient use of evaluation resources and would be of little interest to senior managers. However, relevance issues were not the only issues that were perceived to be inapplicable for specific evaluations; for example, the applicability of each core issue could depend on factors related to the program and its environment. Case studies indicated that in both large and small organizations, the core issues requirement may have led to less useful evaluations that were more complex, time-consuming and resource-intensive than they could have been.

Deputy heads consulted in 2014 called for increased flexibility in applying core evaluation issues, to allow organizations to focus on areas of interest and to produce concise reports to inform decision making. At the same time, deputy heads recognized that increased flexibility would require senior management to become more engaged in evaluation design.

Most stakeholder groups favoured adding flexibility to the requirement for addressing all five core issues. ⁶⁹ Although consultation findings showed that central agencies generally favoured the consistent application of all five core issues across evaluations, they, as well as program managers and evaluators, agreed that flexibility should be considered. Heads of evaluation desired flexibilities that ranged from not addressing a core issue, to minimizing the examination of a core issue, or to addressing one or more core issues less frequently—for example, in every second evaluation of the program or every ten years.

Among the small departments examined in the case studies, the core issues were perceived to increase the burden of evaluations without a corresponding internal demand for the information, leading these organizations to exercise their option to not address them all. ⁷⁰

It was generally noted by stakeholders that section 42.1 of the *Financial Administration Act* calls for a review of relevance and effectiveness every five years, without mandating the five core issues. The added requirement of both the *Policy on Transfer Payments* and the *Policy on Evaluation* for these evaluations to address all five core issues was viewed by evaluation units as imposing an added burden. At the same time, central agency analysts, who are key users of evaluations for advising the Treasury Board on spending proposals such as renewals for programs of grants and contributions, generally supported the need for addressing all five core issues.

The core issues were generally perceived to be very broad in nature. Although no other issue was consistently identified as one that should be added to the existing set of core issues, deputy heads and senior managers sometimes wanted individual evaluations to address additional issues. ⁷¹ As the *Directive on the Evaluation Function* explicitly permits, evaluations often addressed additional issues to meet departmental users' needs. ⁷² For example, 61% of evaluators surveyed indicated that issues other than the five core

issues were addressed in evaluations at least half the time. When evaluators were asked which issues, other than the five core issues, were considered for inclusion in evaluations since 2009, the most frequent issue related to program design and delivery. Consultations with evaluators and central agencies indicated that design and delivery issues had sometimes been addressed under one or both of the two core performance issues; for example, the *Directive on the Evaluation Function*'s description of the core issue "achievement of expected outcomes," explicitly refers to the assessment of program design as part of this issue. Other non-core issues were identified, ⁷³ which stakeholders stated could usually ⁷⁴ or sometimes ⁷⁵ be subsumed within the core issues.

In applying their available evaluation resources, evaluation units sometimes had to choose between addressing a non-core issue to satisfy their deputy head's information needs and complying with the core issues requirement. Consultations indicated that they tended to consider the deputy head's needs to be more important.

As noted in the Implementation Review, although flexibilities existed for addressing core issues with less evaluation effort or for providing a rationale for not addressing an issue, these flexibilities may not have been fully communicated to, or used by, departments. This finding may be explained, in part, by their concerns that Management Accountability Framework assessment ratings would be adversely affected. ⁷⁶ Deputy heads mentioned this rationale in consultations, noting that while further flexibilities may be needed for the core issues requirement, existing flexibilities had not been fully used. Although stakeholders were aware of flexibilities, some perceived them as largely theoretical and saw no real opportunity for discretion when applying the core issues in specific evaluations, and in particular, no opportunity to justify that an issue should not be addressed. According to many stakeholders, the effort required to demonstrate that an evaluation does not need to address certain relevance issues (alignment with federal roles and responsibilities and alignment with federal government priorities) could be equivalent to the effort required to actually evaluate them. Many organizations did, however, calibrate their efforts ⁷⁷ on these relevance questions.

14. Finding: Longstanding inadequacies in the availability and quality of program performance measurement data and incompatibly structured financial data continued to limit evaluators in providing assessments of program effectiveness, efficiency (including cost-effectiveness) and economy. Central agencies and senior managers desired, in particular, more and better information on program efficiency and economy.

Although it was not a specific focus of the evaluation, the impact of performance measurement data on the success of evaluation was a common theme that emerged. The [2009 Fall Report of the Auditor General of Canada](#) noted that many evaluations did not adequately assess program effectiveness because analysis was limited by inadequate performance measurement data. This finding was echoed in the [2013 Spring Report of the Auditor General of Canada](#), which noted that in 14 of 20 evaluations sampled across three departments, weaknesses in program performance measurement data continued to limit evaluators in assessing program effectiveness and often required them to rely on more subjective data or to collect additional data to fill gaps. Although the Implementation Review found that the situation had improved to some extent in 2012–13, about 50% of organizations continued to report challenges related to insufficient performance measurement data. Although the survey of program managers found that 90% of programs had a performance measurement strategy in place, the same survey found that roughly 50% of the indicators were only somewhat useful, not very useful or not at all useful to evaluators. The ongoing nature of this challenge was further reflected in the Secretariat's 2013–14 Capacity Assessment Survey, where 30 of 33 evaluations approved between April and September 2013 experienced significant limitations owing to insufficient performance measurement data—despite a reported 70% rate of implementation of performance measurement strategies. The *Directive on the Evaluation Function*'s requirement for evaluation units to annually produce reports for departmental evaluation

committees on the state of performance measurement helped bring attention to these problems in a number of departments.

Case studies and stakeholder consultations showed a general consensus that despite the core issue requiring evaluations to examine program efficiency and economy, evaluations rarely met the information needs of users in this regard. For example, inadequate financial information limited the ability of evaluators to analyze the cost-effectiveness of programs. Although the available financial data were not inaccurate or incomplete, they were not structured according to the entity being evaluated—that is, resources were not linked to program activities or outcomes. There was concern among stakeholders that the Secretariat's guidance documents had provided too few concrete examples of how to assess efficiency and economy and did not address the prevailing structural challenges of financial data.

3.3 Approaches to Measuring Policy Performance (evaluation question 4)

15. Finding: Mechanisms for measuring policy performance tracked the obvious uses of evaluations—those that were direct and more immediate—but did not capture the range of indirect, long-term or more strategic uses, and may not have given a robust perspective on the usefulness of evaluations.

The *1996 May Report of the Auditor General* noted that “there is still no systematic process in place in most departments to objectively assess and demonstrate the value obtained from evaluation. Few departments have such mechanisms.” ⁷⁸ However, the document review revealed some ways in which departments and central agencies now track the use of evaluation.

The 2009 policy requires departments to monitor compliance to ensure effective implementation and requires departmental evaluation committees to ensure follow-up to action plans approved by deputy heads. The composition of departmental evaluation committees, which generally includes assistant deputy ministers and the deputy head as chair, increased senior management attention to the implementation of action plans and may have also contributed to increased evaluation quality ⁷⁹ and led evaluation functions to serve more strategic evaluation uses.

Several lines of evidence showed that departments generally supported the implementation of evaluation recommendations through a more rigorous approach to tracking follow-up on action items in management responses and action plans. ⁸⁰ More rigorous tracking appeared to be influenced by the Secretariat's Management Accountability Framework assessment criteria, which looked for the presence of tracking systems. However, in 2011 the Centre of Excellence for Evaluation had noted that tracking processes relied to a large extent on self-reported evaluation use and may have been based on perceptions of use. The Secretariat indicated that additional indicators would be needed to validate and enhance the monitoring of evaluation use.

Several lines of evidence also showed that the tracking of action items did not measure the entire range of evaluation impacts; for example, several deputy heads consulted in 2011 reported that some uses of evaluation were unpredictable and untraceable. ⁸¹ Some evaluators and program managers ⁸² reported that the tracking systems were mechanistic and that they did not measure the entire range of evaluation impacts.

The document review distinguished between several types of evaluation use, including:

- Process use (defined as program and operational changes that occur not because of evaluation findings or recommendations, but as a result of the evaluation process itself);
- Instrumental use (where evaluation findings directly inform a decision or contribute to solving a problem) for specific program improvements; and

- Knowledge use at a horizontal level (where evaluations broaden thinking about a program or policy over time and beyond the specific program evaluated).

In 2012, the Centre of Excellence for Evaluation acknowledged that departments lacked resources and methods to track evaluation uses other than instrumental use and, in particular, to trace the influence of evaluation findings in policy discussions and program transformations. Evaluators were generally unaware of some evaluation uses, and in the case studies, evaluation teams noted that they had no systematic approach to measuring uses.

However, stakeholder consultations found that in some organizations there were processes in place to gather information about the use of evaluations beyond the implementation of their recommendations. For example, in some departments there were routine consultations with directors general three months after evaluations, and in others, there were annual reports on evaluation use. Some evaluation units ⁸³ were believed to survey users following evaluations.

The approaches that the Secretariat used for monitoring policy performance included the Capacity Assessment Survey, Management Accountability Framework ratings (up to 2013–14), reviews of departmental evaluation plans, assessments of evaluation report quality, ad hoc consultations and the Annual Report on the Evaluation Function, which brings together many of these information sources. Case studies showed that there was no direct alignment between the utility of evaluations in departments and the Secretariat's assessments of evaluation quality. This finding may be explained by the fact that the evaluation report assessments focused on the report alone (including methodology and structural elements, such as whether the report listed limitations), rather than on the entire evaluation process and the dissemination of findings. In contrast, the case studies demonstrated that departments often used evaluation findings that had not been documented in evaluation reports, or they applied other learning that arose through the evaluation process. The limitations of the approaches used by the Secretariat to measure policy performance in terms of the use and utility of evaluations for various purposes were also documented in the [2012 Annual Report on the Health of the Evaluation Function](#).

3.4 Other Findings

16. Finding: The requirements of the *Policy on Evaluation* and those of other forms of oversight and review, such as internal audit, created some overlap and burden.

There were real or perceived overlaps between the audit and evaluation functions that led to confusion for some stakeholders and to potential burden on programs. In case study consultations, some departments and, in particular, some program managers were fatigued by the increase in evaluations and other forms of oversight. Some respondents suggested, for example, that evaluations be conducted every five years only if no other form of assessment (for example, audit or review) had been conducted. It was noted that some respondents, including program managers, senior managers and departmental evaluation committee members, did not appear to understand the distinctions between evaluation and audit—that is, the distinction between evaluation's focus on program relevance and performance and audit's focus on compliance, control and management performance.

However, deputy heads who were consulted generally understood that the two functions provided different analytic focuses and distinct values. For example, audits examine issues such as controls, probity and compliance, whereas evaluations focus on performance (effectiveness, cost-effectiveness and relevance). However, deputy heads did comment on a specific overlap related to performance audits (which, until 2004, were called value-for-money audits by the Office of the Auditor General), as they indicated that the scope of performance audits sometimes resembled evaluations. At the same time, they noted that some evaluations

examined traditional audit issues. These views were reflected to a lesser extent by program managers, central agency analysts and evaluators. ⁸⁴ Departmental stakeholders made particular mention of transfer payment programs that require the recipient to commission performance audits, which, depending on how they are executed, created potential duplication with evaluations and added burden on recipients and clients. In these cases, consulted departments saw a need for greater flexibility to decide whether an evaluation or an audit would be the most appropriate oversight tool for a program.

The Implementation Review noted that evaluation and audit functions were increasingly co-located or led by one individual playing the role of head of evaluation and chief audit executive, but the effects of this phenomenon were unclear. Some heads of evaluation noted the potential for better coordination of audit and evaluation planning; others indicated that there could be challenges to the required independence of the chief audit executive role and possible limitations for the career path of evaluation executives, as the qualifications required of chief audit executives would require an evaluation executive to become a certified auditor.

4.0 Conclusions

The 2009 *Policy on Evaluation* helped the government-wide evaluation function play a more prominent role in supporting the Expenditure Management System, by making evaluation information more systematically available. This provided departments with a predictable stream of performance information to use for expenditure management and for other uses such as program and policy improvement, accountability and public reporting, including for spending that had not been previously evaluated. Strong engagement from deputy heads and senior management in the governance of the evaluation function created conditions that supported the utility of evaluations, and the evaluation needs of deputy heads, senior managers and central agencies were well served. In some cases, but not systematically across departments, evaluation functions produced horizontal analyses that contributed to useful cross-program learning, informing improvements both to the program evaluated and to other programs and to the organization as a whole. However, in assessing program effectiveness, efficiency and economy in evaluations, departmental functions were often limited by inadequacies in the availability and quality of performance measurement data and by incompatibly structured financial data, which left users (in particular central agencies) wanting more and better information.

Although the comprehensive coverage model appropriately reflected the 2009 policy's objective to serve multiple evaluation users and purposes, the standardized requirements for coverage of all direct program spending every five years and for examination of core issues in all evaluations did not always produce evaluations that closely aligned with user needs. There was a general belief across stakeholders that all government spending should be evaluated periodically, but there was also a widely held view that the potential for individual evaluations to be used should influence their conduct. Further, the policy requirements for evaluation timing and focus did not leave sufficient flexibility for departmental evaluation functions to fully reflect the needs of users when planning evaluations, or to respond to emerging priorities. Evaluation needs were found to vary among different user groups—in particular, the needs of central agencies and departments were different. However, to fulfill coverage requirements within their resource constraints, departments sometimes chose evaluation strategies (for example, clustering programs for evaluation purposes) that were economical but ultimately served a narrower range of users' needs. The lack of flexibility in the coverage and frequency requirements also made it challenging for departments to coordinate evaluation planning with other oversight functions in order to maximize the usefulness of evaluations and minimize program burden.

5.0 Recommendations

The evaluation recommends that when developing a renewed *Policy on Evaluation* for approval by the Treasury Board, the Treasury Board of Canada Secretariat should:

1. Reaffirm and build on the 2009 *Policy on Evaluation*'s requirements for the governance and leadership of departmental evaluation functions, which demonstrated positive influences on evaluation use in departments.
2. Add flexibility to the core requirements of the 2009 *Policy on Evaluation* and require departments to identify and consider the needs of the range of evaluation user groups when determining how to periodically evaluate organizational spending (including the scope of programming or spending examined in individual evaluations), the timing of individual evaluations, and the issues to examine in individual evaluations.
3. Work with stakeholders in departments and central agencies to establish criteria to guide departmental planning processes so that all organizational spending is considered for evaluation according to the core issues; that the needs of the range of key evaluation users, both within and outside the department, are understood and used to drive planning decisions; that the planned activities of other oversight functions are taken into account; and that the rationale for choices related to evaluation coverage and to the scope, timing and issues addressed in individual evaluations is transparent in departmental evaluation plans.
4. Engage the Secretariat's policy centres that guide departments in the collection and structuring of performance measurement data and financial management data in order to develop an integrated approach to better support departmental evaluation functions in assessing program effectiveness, efficiency and economy.
5. Promote practices, within the Secretariat and departments, for undertaking regular, systematic cross-cutting analyses on a broad range of completed evaluations and using these analyses to support organizational learning and strategic decision making across programs and organizations. In this regard, the Treasury Board of Canada Secretariat should facilitate government-wide sharing of good practices for conducting and using cross-cutting analyses.

Appendix A: Evolution of Evaluation in the Federal Government and the Context for Policy Renewal in 2009

Evaluation was officially introduced in the federal government in the late 1970s to help improve management practices and controls. The 1977 *Evaluation Policy* ⁸⁵ mandated that evaluation be a component of each organization's management and that all programs be evaluated periodically, every three to five years. The policy recognized evaluation as a deputy head's managerial responsibility. Deputy heads were to use evaluation findings and recommendations about program effectiveness and efficiency to inform decisions on management and resourcing, to be accountable for their programs, and to provide quality advice to ministers.

When renewed in 1992, the policy ⁸⁶ recommended a six-year cycle for evaluating the continued relevance, success and cost-effectiveness of federal programs, but noted that when there was no priority need for this performance information or when it would require excessive time or resources, no evaluation should be conducted. The policy called for evaluation criteria to be established for all programs, as the means by which performance could be judged. Evaluations were to be used to reconfirm, improve or discontinue programs, and departmental evaluation planning was expected to respond to evaluation issues that reflected concerns of the Treasury Board or other Cabinet committees.

In 1994, a *Review Policy* ⁸⁷ brought together performance measurement and review requirements under one umbrella and included internal audit and evaluation. It emphasized the responsibility of line managers to

demonstrate performance and to manage for results, and aimed to promote collaboration between managers and reviewers.

A study of the evaluation function in 2004 ⁸⁸ examined the *Review Policy* and identified the need for a clear distinction between internal audit and evaluation, to better serve the needs of managers.

The 2001 *Evaluation Policy*, including its “Evaluation Standards for the Government of Canada,” separated the evaluation function from the internal audit function and extended the scope of evaluation planning to include programs, policies and initiatives. The policy focused on results-based management and aimed to embed the discipline of evaluation into management practice. It called on departments to establish strategically focused evaluation plans based on assessments of risk, departmental and whole-of-government priorities, and reporting requirements. The standards asked evaluators to consider the full range of issues when planning evaluations, including program relevance, success and cost-effectiveness, and to address issues needed for accountability reporting.

The 2009 *Policy on Evaluation* establishes a more prominent role for the evaluation function in supporting the Expenditure Management System. Policy requirements for comprehensive evaluation coverage every five years and for evaluations to systematically assess five core issues pertaining to program relevance and performance are intended to address the growing need for neutral, credible evidence on the value for money of government direct program spending to inform expenditure management decisions, as well as policy and program improvement decisions, Cabinet decision making and public reporting. The policy and the associated directive and standard include measures to ensure evaluation quality, neutrality and use.

Evolution of Policy Requirements for Extent and Frequency of Evaluation Coverage and Evaluation Issues

Over the years that federal evaluation policies have existed, there have been various approaches to the extent and frequency of evaluation coverage, and the evaluation issues examined in evaluations.

Coverage requirements have existed in some manner in all Treasury Board evaluation policies, ranging from ensuring that all programs are evaluated periodically (the 1977 policy) to considering, but not requiring, evaluation of all programs (the 2001 policy), and most recently, the 2009 policy's requirement to evaluate all direct program spending. Although the frequency of evaluation has varied in federal policies, from every three to five years (the 1977 policy), to every six years (the 1992 policy), and now to every five years (the 2009 policy), a requirement for periodic evaluation has always existed.

All federal evaluation policies have specified a set of issues for evaluations to address. As shown in Table 2, the set of issues has been similar since 1992 and has included program relevance, effectiveness and efficiency.

Table 2: Evolution of Evaluation Issues in Government of Canada Evaluation Policies, 1977 to 2009

	1977 Evaluation Policy	1992 Evaluation Policy	1994 Review Policy	2001 Evaluation Policy	2009 Policy on Evaluation
Evaluation Issues	<ul style="list-style-type: none"> • Effectiveness • Efficiency 	<ul style="list-style-type: none"> • Relevance • Success • Cost-effectiveness 	<ul style="list-style-type: none"> • Relevance • Success • Cost-effectiveness 	<ul style="list-style-type: none"> • Relevance • Success • Cost-effectiveness 	<ul style="list-style-type: none"> • Relevance issue 1: Continued need for program • Relevance issue 2:

					<p>Alignment with government priorities</p> <ul style="list-style-type: none"> • Relevance issue 3: Alignment with federal roles and responsibilities • Achievement of expected outcomes • Resource utilization (demonstration of efficiency and economy)
--	--	--	--	--	--

In comparing the evaluation issues of the 2009 policy with those of the 2001 policy, a key difference is that “relevance” is divided into three issues. This change aligned the 2009 issues with the key objectives for strategic reviews of federal programming, which were underway at the time the policy was renewed. In 2009, the cost-effectiveness issue was also recast as an examination of program resource utilization, which was intended to give evaluators more flexibility in selecting assessment approaches. [89](#)

A notable change in 2009 was that core evaluation issues were no longer discretionary. Under the 2009 policy, all five core issues must be addressed for evaluations to meet coverage requirements; however, departments have the flexibility to determine the evaluation approach and level of effort applied. In contrast, the 2001 policy indicated that “the full range of evaluation issues *should be considered* [emphasis added] at the planning stage of an evaluation...” and that “evaluators *should* [emphasis added] address issues that are needed for accountability reporting, including those involving key performance expectations....”

Context of the Federal Evaluation Function Leading Up To Policy Renewal in 2009

In the years before policy renewal in 2009, several factors increased the demand for credible information about program relevance, effectiveness, efficiency and economy. The major factors were changes to key legislation and to the Expenditure Management System.

A 2006 Legislated Requirement for Evaluation

In June 2006, the President of the Treasury Board commissioned an independent blue ribbon panel, “to recommend measures to make the delivery of grants and contributions programs more efficient while ensuring greater accountability.” Following the release of the panel's report, [90](#) the *Federal Accountability Act* of December 2006 amended the *Financial Administration Act* to require that all ongoing programs of grants and contributions be reviewed for relevance and effectiveness every five years. The 2008 *Policy on Transfer Payments* and the 2009 *Policy on Evaluation* later defined these reviews as evaluations, and the 2009 *Policy on Evaluation* reflected the legal requirement in its own coverage requirements. Before 2006, the requirement for reviewing or evaluating ongoing programs of grants and contributions was contained only in the *Policy on Transfer Payments*, where it continues to appear today.

A Shift in Emphasis for Evaluation to Support Expenditure Management

The renewal of the Expenditure Management System in 2007 placed greater emphasis on using evaluation as an input to expenditure decisions. In accordance with Budget 2006 commitments, the renewed Expenditure Management System is based on the following key principles:

- Government programs should focus on results and value for money;
- Government programs must be consistent with federal responsibilities; and
- Programs that no longer serve the purpose for which they were created should be eliminated.

The system's renewal addressed the November 2006 Auditor General's recommendations that expenditure decisions be anchored by reliable information on program performance. In February 2008, the Standing Committee on Public Accounts ⁹¹ recommended that “the Treasury Board Secretariat reinforce the importance of evaluation by adding program evaluation as a key requirement in the Expenditure Management System.”

In the renewed Expenditure Management System, evaluation is positioned as an important source of neutral, credible evidence about program value for money, to provide support to expenditure decisions for each of the three pillars of the system, as follows:

Managing for Results:

Evaluations are used by departments on an ongoing basis to manage for results—that is, to determine whether programs are achieving expected results and to inform decisions about continuing, amending or terminating program spending.

Upfront Discipline:

Evaluation evidence is used in new spending proposals (such as in the Memorandum to Cabinet process) to help compare proposed spending with existing or past program results.

Ongoing Assessment:

Evaluations provide input to spending reviews (comprehensive or targeted) to support analyses of whether programs are effective and efficient, are focused on results, are providing value for taxpayers' money, and are aligned with government priorities.

Strategic and Other Spending Reviews

Before and after the renewal of the *Policy on Evaluation* in 2009, spending reviews increased the demand for evaluations. Each year from 2007–08 to 2010–11, subsets of departments participated in strategic reviews, led by the Treasury Board of Canada Secretariat, that examined all federal direct program spending over the complete four-year period. The strategic reviews used departmental Program Alignment Architectures as the organizing framework and analyzed spending according to issues such as relevance, alignment with government priorities, and effectiveness and efficiency. Following this period, in 2011–12, all departments were engaged in a comprehensive strategic and operating review.

External Audits of Evaluation Policy in the Government of Canada

In 1993 the Auditor General reported ⁹² that the strength of the evaluation function and the number of evaluations were declining, and that only about one quarter of government spending from 1985–86 to 1991–92 was evaluated, far short of expectations that all programs be evaluated over five years. Across all departments in 1991–92, \$28.5 million was spent on evaluation.

In 1996 the Auditor General reported ⁹³ that evaluation coverage had improved but that some programs over \$1 billion had not been evaluated and that evaluations were normally too focused on smaller program components and lower-level issues: the choices that departments made reflected their interests and priorities but did not necessarily produce information on program effectiveness to support accountability and government decision making.

In 2000 the Auditor General ⁹⁴ found that the evaluation function had regressed and that funding reductions undermined evaluation capacity.

In 2009 the Auditor General underscored ⁹⁵ that “a vital purpose is served when effectiveness evaluation informs the important decisions that Canadians are facing.” By examining a sample of departmental evaluations conducted between 2004 and 2009, which was a period governed by the 2001 *Evaluation Policy*, the Auditor General found that 5% to 13 % of spending was evaluated annually and concluded that departments' low evaluation coverage and inadequate collection of performance measurement data meant that needs for information about program effectiveness were not being adequately met. The Auditor General noted that although evaluation funding and staff had increased during this period, departments found it challenging to meet evaluation requirements.

In a follow-up audit in 2013, the Auditor General concluded, “Implementation of the 2009 *Policy on Evaluation* has supported improvements in a number of areas. However, significant weaknesses continue to limit the contribution of evaluation to decision making in the government.” ⁹⁶ Even though three quarters of large organizations planned to achieve comprehensive five-year coverage by 2017, the Auditor General reported unsatisfactory progress on coverage. The audit found that departments had made progress since 2009 in generating ongoing performance information, but that program evaluators still noted constraints in being able to address program effectiveness owing to limited availability of ongoing performance information. As a result, departments were making decisions about programs and related expenditures with incomplete information about their effectiveness.

The 2013 audit also found that departments were concerned about the policy requirements for evaluating all programs every five years and for addressing the full range of evaluation issues in all evaluations. Departments indicated that although they had the capacity to achieve comprehensive coverage over five years, the combined requirements for coverage and core issues limited the extent to which they could put their evaluation resources to best use. The Auditor General indicated that the Treasury Board of Canada Secretariat should consider these concerns when evaluating the *Policy on Evaluation* in 2013–14.

Appendix B: Implementation Review of the 2009 *Policy on Evaluation*

After the introduction of the 2009 *Policy on Evaluation*, the Treasury Board of Canada Secretariat continuously monitored and reported on its implementation. To identify issues, the Secretariat completed an Implementation Review in 2013 that examined the policy's four-year transition period before five-year comprehensive evaluation coverage was fully implemented. The review involved broad consultations with over 140 stakeholders at all levels, in departmental evaluation functions, program areas and central agencies. The review team included analysts from the Secretariat's Centre of Excellence for Evaluation as well as consultants from Hickling Arthurs Low (HAL) Corporation, and was supported by a review advisor, Dr. William Trochim of Cornell University, who provided advice on the overall approach, methodology and planning for the review. In addition, two advisory committees provided feedback on the review plans and draft reports: one included departmental heads of evaluation, and the other included central agency representatives.

Taken together, the Implementation Review and the Secretariat's Annual Reports on the Health of the Evaluation Function from 2010 to 2012 showed that departments had made solid progress during the policy's four-year transition period in the following areas:

- In general, departments had implemented structures, roles and responsibilities for governing the evaluation function and planning its activities, and demonstrated greater engagement by senior management in departmental evaluation committees.
- Departments had built capacity and progressed toward full implementation of the policy's requirements starting in April 2013:
 - The number of full-time equivalents working in the evaluation function across government had increased from 409 in 2007–08 to 500 ⁹⁷ in 2011–12, and financial resources had remained relatively stable, in the \$60 million range;
 - There was a notable increase in evaluation coverage by large departments, compared with pre-policy levels: the average annual coverage of direct program spending increased from 6.5% in 2007–08 to 16.8% in 2011–12;
 - Evaluations of more highly aggregated programming were common—a prevalent strategy for increasing coverage without increasing evaluation resources to the same extent; and
 - Many large departments had produced plans for comprehensive coverage before 2013–14, even though the policy permitted risk-based planning of coverage before this date.
- Departments had increasingly used evaluation to support decision making, such as for informing spending review proposals and preparing Treasury Board submissions and Memoranda to Cabinet.

Implementation of Policy Requirements Related to Leadership, Governance and Planning

In general, departments had implemented policy requirements related to the roles and structures for leading and governing departmental evaluation functions, and tools for planning evaluations.

Heads of Evaluation

The Implementation Review found that as departments implemented the policy, there was a shift toward designating heads of evaluation at higher executive levels, enabling them to play more strategic advisory roles in departmental decision making. Almost two thirds of heads of evaluation (64%) were designated at the EX-3 and EX-4 levels in 2012–13, compared with less than one third (30%) in 2009–10.

The review found that over the same four-year period, the pairing of the evaluation function with other functions became a common practice in both large and small organizations, as did the practice of combining the head of evaluation role with other leadership roles. In 2012–13, roughly three quarters of heads of evaluation fulfilled leadership roles in two or more other functions; in particular, 61% of heads of evaluation fulfilled the role of chief audit executive, compared with 39% in 2009–10. The prevalence of paired audit and evaluation units also increased, from 41% of departments in 2009–10 to 67% of departments in 2011–12. This trend toward pairing evaluation with other functions and combining the head of evaluation role with other roles appeared to have been driven primarily by the policy requirement for heads of evaluation to have unencumbered access to the deputy head. This trend was possibly enhanced by organizational restructuring following government-wide cost-containment exercises.

Departmental Evaluation Committees

The Implementation Review found that by the end of the policy's transition period, all large departments had established departmental evaluation committees, the large majority being chaired by deputy heads. Increasingly, evaluation committee members were senior decision makers representing all or most organizational divisions. In most cases, committee members were senior executives who were also members of the senior executive committee. In one third of departments in 2011–12, the membership composition of the departmental evaluation committee matched that of the senior executive committee. Departmental evaluation committees were also increasingly involved in activities such as tracking individual evaluation recommendations and advising on resources required by the function.

Departmental Evaluation Plans

The 2009 policy requires that deputy heads annually approve a rolling five-year departmental evaluation plan that aligns with and supports the Management, Resources and Results Structure, that supports the Expenditure Management System, and that evaluates all ongoing programs of grants and contributions, as required by section 42.1 of the *Financial Administration Act*. The evaluation plan is a vehicle for communicating within departments and with the Treasury Board of Canada Secretariat, especially with analysts involved in expenditure management processes.

The Implementation Review found that since 2009, more than 90% of large departments and agencies had annually submitted plans to the Secretariat. A majority of departments developed their plans according to the Secretariat's guidance, by broadly consulting with program areas and discussing performance measurement needs for supporting evaluation.

Capacity for Fully Implementing Policy Requirements

Resource Allocation

At the time the 2009 policy was introduced, the Treasury Board of Canada Secretariat had projected that departments would need to increase investment in the evaluation function to achieve and sustain comprehensive evaluation coverage every five years. Despite a temporary increase in resources in 2009–10, the Secretariat's monitoring showed that government-wide financial resources for the function were stable, at about \$60 million annually until 2011–12.

Although financial resources for the evaluation function were relatively stable during the policy's transition period, the number of full-time equivalents dedicated to the function in 2011–12 was somewhat higher (500 ⁹⁸) than in 2009–10 (474 ⁹⁹). This increase in human resources appeared to be achieved by decreasing budgets for professional services and reallocating these funds to salaries.

Human Resources Capacity Building

After the Secretariat's regular engagement with departmental evaluation functions up to 2011–12 and a survey of federal evaluators to determine their professional development needs, it was concluded that introductory-level evaluation training was no longer a high-priority need. However, when heads of evaluation were consulted a year later for the Implementation Review, they indicated that some evaluators who had recently joined the function lacked evaluation expertise, causing novice-level training to re-emerge as a short-term need.

In addition, a number of heads of evaluation and directors of evaluation felt that the need for more experienced evaluators had increased under the 2009 policy because departments were choosing more

complex evaluation designs. Specifically, they noted that the clustering of programs for evaluation purposes had increased the demand for expertise and experience typically held by senior evaluators.

Strategies Used to Implement Coverage Requirements

The Implementation Review found that departments used one or more of the following strategies to expand their evaluation coverage within their budgeted resources.

Clustering Programs for Evaluation Purposes

The 2009 policy allows departments to group or divide direct program spending for the purposes of evaluation, as appropriate to decision-making needs. The Implementation Review found that this strategy often resulted in evaluations that covered the “practical” units of programming, and expenditure management (for example, programs defined through Treasury Board submissions) but did not necessarily match programs as defined in Program Alignment Architectures.

According to a review of 28 departmental evaluation plans submitted to the Secretariat in 2012–13, 81% of departments used clustering strategies to achieve evaluation coverage. By clustering, departments conducted fewer evaluations, with greater economies of scale—for example, by reducing the effort required for planning, conducting or procuring several smaller evaluations. Clustering typically involved grouping low-dollar value programs or programs that had common intended outcomes or objectives, themes or delivery models.

Calibrating Evaluation Effort

The Implementation Review found that departments understood that they could adjust the scope and depth of analysis in evaluations, but that the options and limits for calibrating the evaluation were unclear to them. As a result, departments may not have fully exploited calibration possibilities.

Relying More on Internal Evaluators

To use their resources more efficiently, many departments relied increasingly on internal evaluators to lead and conduct evaluations, using external evaluators only for specific evaluation tasks (for example, data collection) or to provide more capacity when internal capacity was insufficient.

Minimizing Non-Evaluation Activities

To focus resources on meeting coverage requirements, a number of departments chose to minimize non-evaluation activities, such as other types of research and assistance to programs for the development of performance measurement strategies.

Performance Measurement to Support Evaluation

The 2009 *Policy on Evaluation* requires program managers to develop and implement performance measurement strategies, which support future evaluations and also the ongoing management of their programs. Program managers consult with heads of evaluation to ensure that their strategies will produce data that meet evaluation needs. In addition, the policy requires heads of evaluation to prepare an annual report on the state of performance measurement for the departmental evaluation committee.

The Implementation Review found that from 2009–10 to 2011–12, the proportion of evaluations that were supported by performance measurement data rose from 62% to 78%. ¹⁰⁰ However, the percentage of evaluation reports that indicated that data quality was sufficient or partially sufficient for evaluation needs did not increase to the same extent (from 49% in 2009–10 to 52% in 2011–12). Where performance

measurement data were insufficient, evaluators often could not do meaningful analyses of program effectiveness.

A large majority of review informants said that the annual report on the state of performance measurement drove discussions between evaluation units and program areas and ultimately led to the development and implementation of performance measurement strategies. Further, heads of evaluation and directors of evaluation observed that the report increased the attention paid by senior management to performance measurement and areas where improvements were needed.

Implementation Progress in Small Departments and Agencies

The Implementation Review drew tentative findings about policy implementation in small departments and agencies because of the limited number that were consulted and the discontinuous monitoring of these organizations by the Treasury Board of Canada Secretariat.

In contrast to the 2001 *Evaluation Policy*, the 2009 *Policy on Evaluation* no longer requires small organizations to establish departmental evaluation committees or to develop departmental evaluation plans. For small organizations, the 2009 policy deferred the requirement to comprehensively evaluate all direct program spending.

However, the 2009 policy requires deputy heads of small organizations to:

- Designate a head of evaluation having unencumbered access to the deputy head;
- Approve evaluation reports, management responses and action plans and make them publicly available;
- Ensure that all ongoing programs of grants and contributions are evaluated every five years, as required by section 42.1 of the *Financial Administration Act*; and
- Ensure that other direct program spending is evaluated as appropriate to the needs of the department.

The Implementation Review observed that in some small organizations the deferral of key policy requirements led to an erosion of evaluation functions and the dismantling of some evaluation infrastructure. However, other small organizations retained some evaluation infrastructure and processes and some chose to maintain functional features required by the policy only in large organizations. For example, several small organizations maintained their departmental evaluation committees or integrated their responsibilities into other governance committees, such as their executive committees.

The Implementation Review found that small organizations had designated heads of evaluation and that 80% of them had unencumbered access to their deputy heads. The seniority level of heads of evaluation was slightly lower in small organizations than in large ones. All heads of evaluation in small organizations also played leadership roles in other functions (for example, internal audit, performance measurement or risk-management), and 90% of evaluation functions were co-located with another function, such as strategic planning, performance measurement or audit.

Challenges of Policy Implementation

While the expansion of evaluation coverage coincided with greater use of evaluation and other benefits, it presented challenges for some departments. For example:

- Some found it challenging to allocate adequate resources to the evaluation function to support comprehensive coverage over five years;
- Some may not have known how to apply the policy's flexibilities for calibrating evaluations and clustering program spending to achieve more cost-effective coverage. Furthermore, the Secretariat's approach to

rating evaluation quality through the Management Accountability Framework assessment process may have made departments less likely to experiment with calibrated approaches; and

- Departments generally felt that the five-year comprehensive coverage requirement, and to an extent the *Financial Administration Act* requirement, limited their responsiveness to emerging evaluation priorities and the information needs of deputy heads.

Although the universal application of the five core issues supported the utility of evaluations across the spectrum of uses and users targeted by the policy, it presented challenges for some departments. Some departments wanted more flexibility in applying the issues, as they felt that not all issues were necessary in all evaluations, nor a good use of evaluation resources. Given their limited evaluation resources, some departments were of the view that requiring core issues to be universally applied increased the challenge of achieving five-year comprehensive coverage.

Although performance measurement improved somewhat during the first years of policy implementation, evaluation functions still found the data insufficient to fully support evaluation. Weaknesses in the availability and quality of performance measurement data presented challenges to the efficient use of evaluation resources, as evaluators took steps to compensate for a lack of data in order to assess program value for money.

Appendix C: Purpose of the Evaluation of the 2009 *Policy on Evaluation*, Methodology and Governance Committees for the Evaluation

Purpose of the Evaluation

This evaluation will inform the Treasury Board of Canada Secretariat as it fulfills its responsibilities for policy development and for leading the government-wide evaluation function. The evidence collected during the evaluation will not be used to assess the performance of individual departments in relation to the *Policy on Evaluation*.

This evaluation also mitigates the risks of the *Policy on Evaluation* in not achieving its expected results, “which are to make credible, timely and neutral information on the ongoing relevance and performance of direct program spending available to Ministers, central agencies and deputy heads and used to support evidence-based decision making on policy, expenditure management and program improvements and available to Parliament and Canadians to support government accountability for results achieved by policies and programs.” The complexity of the evaluation was commensurate with the importance of the policy achieving its expected results.

Methodology

Case Studies

Policy Performance Case Studies

External consultants conducted a series of departmental case studies and performed qualitative analyses to draw links between policy implementation and evaluation utilization. A representative sample of 10 departments and agencies was selected. Twenty-eight evaluations that were conducted under the 2009 policy were studied, which represented up to three evaluations per department. In all, 86 key informant interviews

were conducted—between 2 and 14 per case—with heads and directors of evaluation, evaluation team members, managers of evaluated programs, departmental evaluation committee members and central agency officials. Case studies also involved document reviews. A case report was prepared for each organization, synthesizing information from all data sources, and a report on the analysis of all case studies was prepared to synthesize findings across all organizations.

Policy Application Case Studies

To assess the relevance of the *Policy on Evaluation* to various categories of program spending and identify opportunities for flexibly applying key policy requirements, the internal evaluation team conducted case studies of six categories of program spending. The case studies qualitatively analyzed the application of the policy and its key requirements (comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues) to the following six spending categories:

- Assessed contributions to international organizations;
- Endowment funding;
- Programs with a requirement for recipient-commissioned independent evaluations;
- Low-risk programs;
- Programs with a long horizon to achievement of results; and
- Other programs identified by departments as challenging for policy application.

Nine large organizations submitted a total of 22 examples of programs and challenges associated with evaluating them under the policy. Two additional examples were selected by the internal evaluation team. Relevant documents and literature identified by organizations were reviewed, including existing evaluation reports, websites, legislation, funding agreements (or excerpts) as well as other Treasury Board policies and legislation. Subsequently, across the 24 programs, 39 consultations were conducted across 24 departments with a total of 37 departmental representatives, including 14 evaluation professionals and 23 programs managers or others. Consultations were also held with 35 heads of evaluation, or their delegates, in small group settings. These consultations studied the challenges and impacts of addressing the policy requirements and any adjustments made in applying the requirements. Eight consultations were conducted with 11 central agency representatives to get their perspectives on the use and utility of evaluations conducted under the policy, as well as on challenges related to evaluation utility.

Stakeholder Consultations

Stakeholder consultations were conducted as part of ongoing policy dialogue, as well as for informing the evaluation of the *Policy on Evaluation* and the review of the *Policy on Management, Resources and Results Structures*. The Deputy Assistant Secretary of the Expenditure Management Sector at the Treasury Board of Canada Secretariat conducted semi-structured interviews with 15 deputy heads, associate deputy heads and assistant deputy heads from both large and small departments, and with six other key informants, including senior officials from central agencies and former federal public servants.

Surveys

Survey of Program Managers

A survey of departmental program managers was conducted, primarily to inform findings on the impacts of the key policy requirements on the use and utility of evaluations as well as the impacts of other internal and external factors. The online survey was administered using the Secretariat's centrally coordinated survey software. A non-probability sample of 514 program managers was selected from a sampling frame of 707

program managers who were identified by federal departments. A total of 115 responses were received, for a response rate of 22%. Of the 115 respondents, 48 (42%) had managed a program that had been evaluated under the 2001 *Evaluation Policy* and 99 (86%) had managed a program that had been evaluated under the 2009 *Policy on Evaluation*. Many of the survey questions were only asked of those in the latter group. For the survey, program managers were defined as those responsible for managing the units (programs) identified for evaluation in the departmental evaluation plan. The survey results were not weighted owing to a lack of population data.

Survey of Evaluation Managers and Evaluators

A survey of departmental evaluation managers and evaluators was conducted, primarily to inform findings on the impacts of the key policy requirements on the use and utility of evaluations as well as the impacts of other internal and external factors. The online survey was administered using the Secretariat's centrally coordinated survey software. The sampling frame included 392 evaluation managers and evaluators who were identified by federal departments, and all were invited to participate in the survey. A total of 153 responses were received, for a response rate of 39%. Of the 153 respondents, 89 (58%) had been working in the federal evaluation function since before the 2009 policy came into effect. For analysis purposes, data were weighted by position type (evaluation manager, evaluator), by organization size, by the presence of programs of grants and contributions in the department, and by sector. Comparisons were conducted according to sector and to the presence of programs of grants and contributions; however, few systematic differences were found.

Data Analysis

External consultants used SPSS software to perform descriptive and inferential statistical analyses on policy monitoring data previously collected by the Centre of Excellence for Evaluation as well as on the data from the surveys of program managers, evaluation managers and evaluators. Monitoring data included the following:

- Data from annual Capacity Assessment Surveys of departmental evaluation functions, from 2004–05 through 2013–14; and
- Management Accountability Framework assessment ratings of departmental evaluation functions from 2007–08 to 2011–12, including overall ratings and ratings for each of four sub-criteria (Quality of Evaluation Reports, Governance and Support of the Evaluation Function, Evaluation Coverage, and Use of Evaluation), along with overall evaluation ratings for 2006–07.

The focus of the Capacity Assessment Survey changed annually, and the Management Accountability Framework indicators evolved over the years. To conduct longitudinal analyses to show changes that were potentially attributable to the policy, only stable indicators from these two data sources were used.

Open-ended data from the surveys of program managers and the federal evaluation community were coded by the internal evaluation team, and then the coded data were transferred to the external team.

Process Mapping

A process map (see [Appendix D](#)) was developed by the external consultants, through the review of documents and case study data. The map provided an overview of how the evaluation function operates in departments, including processes for planning, conducting and using evaluations.

Document Review

The internal evaluation team conducted a document review to inform questions on the appropriateness of comprehensive coverage and core issues, on the approaches used to measure the performance of the policy, on baseline results for policy outcomes, and on factors affecting the achievement of outcomes. Approximately 60 internal and external documents were reviewed, including Auditor General reports to Parliament, reports of the Standing Committee on Public Accounts, publications of the Treasury Board of Canada Secretariat, and other documents.

Literature Review

The internal evaluation team conducted a literature review to compare and contrast the evaluation policies and practices of other jurisdictions with those of Canada, and in particular, those related to evaluation coverage and frequency and to the core issues addressed in evaluations. The review synthesized the most recently available literature of the following types:

- Official publications from governments (national and sub-national level) and international organizations (agency level), including evaluation policies, guidance documents, evaluation plans, a government constitution, legislation, information reports, evaluation standards, competencies, glossaries and other reports;
- Web pages;
- Academic articles and working papers; and
- Information from presentations given to the Treasury Board of Canada Secretariat.

A sample of nine countries and three international organizations was selected according to several criteria as described below:

- Relevance and comparability to the Canadian context: the United States, the United Kingdom and Australia;
- Comparability of evaluation activity levels, status of evaluation policies, or other reasons: Switzerland, Japan and India—as informed by a 2013 EvalPartners' report, [*Mapping the Status of National Evaluation Policies*](#);
- For continuity, countries examined during the 2013 Implementation Review of the *Policy on Evaluation*: South Africa, Mexico Spain, the United States, the United Kingdom and Australia;
- Evaluation agencies or groups from three international organizations with established evaluation policies or guidance: the United Nations Evaluation Group, the Development Assistance Committee of the Organisation for Economic Co-operation and Development, and the World Bank Independent Evaluation Group; and
- Availability and reliability of online sources and documents in English and French.

Governance Committees for the Evaluation of the 2009 *Policy on Evaluation*

To provide continuity, the Heads of Evaluation Advisory Committee (HEAC) and the Central Agency Advisory Committee (CAAC) that had governed the 2013 Implementation Review continued their roles for the evaluation of the 2009 policy, and members were added to each committee to ensure adequate representation. HEAC membership reflected a range of organization types (for example, large and small organizations in various sectors and with various types of spending (for example, grants and contributions spending), while CAAC membership included the Privy Council Office, the Department of Finance Canada, and the Treasury Board of Canada Secretariat, including its program and policy sectors. The committee's work was governed by terms of reference. The list of committee members is shown in Table 3.

Table 3. Membership of Advisory Committees for the Evaluation of the *Policy on Evaluation*

Heads of Evaluation Advisory Committee	Organization
Shelley Borys, Director General, Evaluation	Public Health Agency of Canada Health Canada
Linda Anglin, Chief Audit and Evaluation Executive	Public Works and Government Services Canada
Susan Morris, Director, Evaluation	Science and Engineering Research Canada Social Sciences and Humanities Research Council of Canada
Stephen Kester, Director, Evaluation	Foreign Affairs, Trade and Development Canada
Courtney Amo, Acting Director, Evaluation and Risk	Atlantic Canada Opportunities Agency
Denis Gorman, Director, Internal Audit and Evaluation	Public Safety Canada
Richard Willan, Chief Audit and Evaluation Executive Marie-Josée Dionne-Hébert, Director, Evaluation Services	Canadian Heritage
Central Agency Advisory Committee	Organization
Renée Lafontaine, Executive Director, International Affairs and Development	Treasury Board of Canada Secretariat, International Affairs, Security and Justice Sector, International Affairs and Development Division
Stephen McClellan, Executive Director, Aboriginal Affairs and Health	Treasury Board of Canada Secretariat, Social and Cultural Sector, Aboriginal Affairs and Health
Catherine Adam, General Director	Department of Finance Canada, Federal-Provincial Relations and Social Policy
Yves Giroux, Director of Operations	Privy Council Office, Liaison Secretariat for Macroeconomic Policy
Mike Milito, Director General, Internal Audit and Evaluation Bureau	Treasury Board of Canada Secretariat, Internal Audit and Evaluation Bureau
Amanda Jane Preece, Executive Director, Results-Based Management Kiran Hanspal, Executive Director, Results-Based Management	Treasury Board of Canada Secretariat, Expenditure Management Sector, Results-Based Management
Nick Wise, Executive Director,	Treasury Board of Canada Secretariat, Priorities and Planning,

Strategic Policy Paule Labbé, Executive Director, Priorities and Planning	MAF and Risk Management Directorate
Sylvain Michaud, Executive Director, Government Accounting Policy and Reporting	Treasury Board of Canada Secretariat, Government Accounting Policy and Reporting

Appendix D: Contribution Theory of the 2009 *Policy on Evaluation*, Generic Evaluation Process Map of the Life Cycle of a Departmental Evaluation, and Logic Model for Implementing the 2009 *Policy on Evaluation*

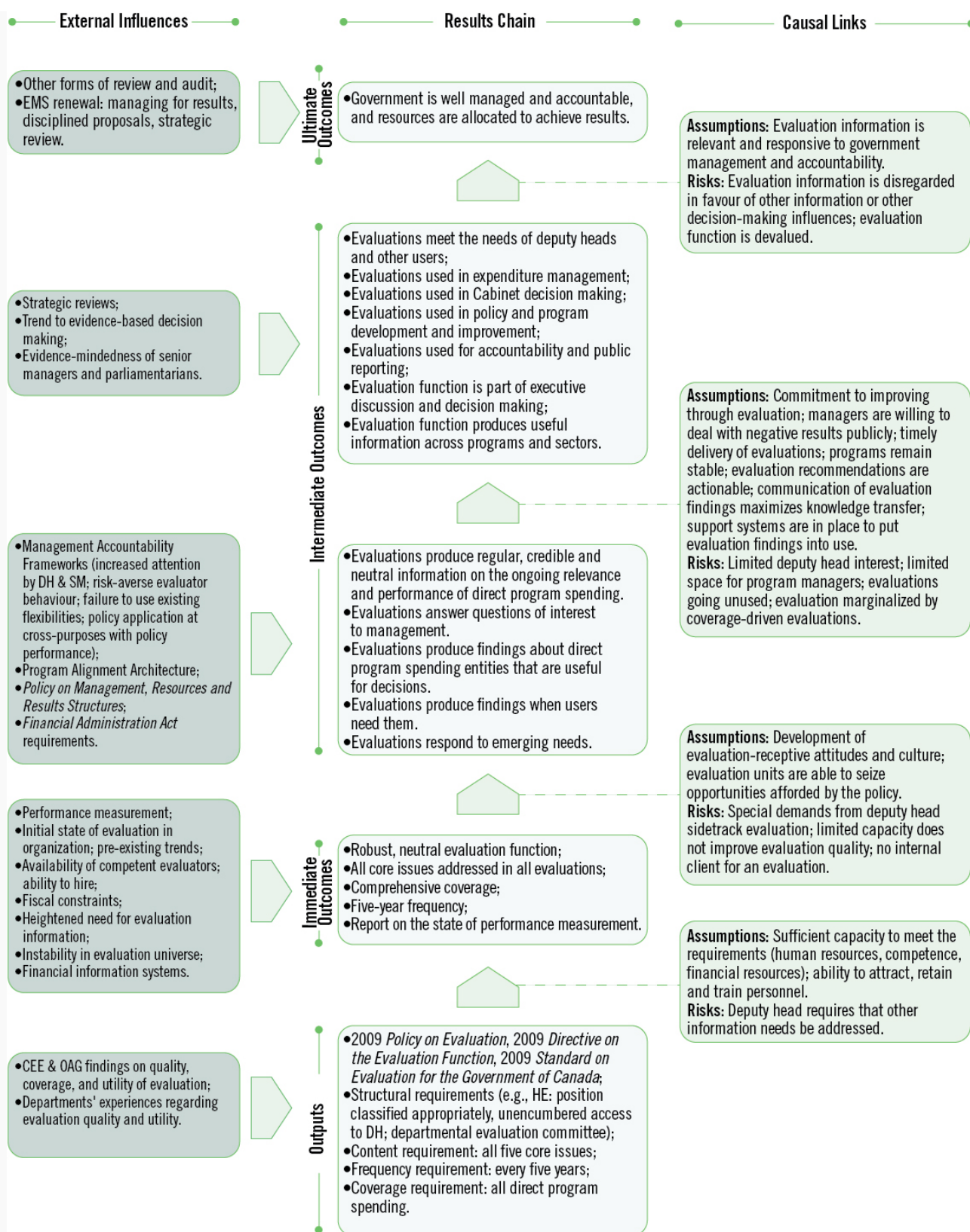
Appendix D presents three figures: the contribution theory of the 2009 *Policy on Evaluation*, a generic evaluation process map of the life cycle of a departmental evaluation, and the logic model for implementing the 2009 *Policy on Evaluation*. A list of abbreviations used in the figures follows.

Abbreviations Used in Figures 7, 8 and 9

Abbreviation	Term
CEE	Centre of Excellence for Evaluation
DEC	Departmental evaluation committee
DEP	Departmental evaluation plan
DH	Deputy head
EMS	Expenditure Management System
FAA	<i>Financial Administration Act</i>
G&C	Grants and contributions
HE	Head of evaluation
MRAP	Management Response and Action Plan
MRRS	Management, Resources and Results Structure
OAG	Office of the Auditor General of Canada
PE	<i>Policy on Evaluation</i>
PMS	Performance measurement strategies
RFP	Request for proposal
SM	Senior managers

TBS	Treasury Board of Canada Secretariat
TR	Terms of reference

Figure 7. Contribution Theory of the 2009 *Policy on Evaluation*



▼ Figure 7 - Text version

The figure shows the theoretical chain of results associated with the 2009 *Policy on Evaluation*, as well as the external influences on the results and the causal links (assumptions and risks) between the four

parts of the results chain, which are (1) outputs, (2) immediate outcomes, (3) intermediate outcomes and (4) ultimate outcomes.

The outputs of the 2009 *Policy on Evaluation* are:

- The policy itself, the 2009 *Directive on the Evaluation Function*, and the 2009 *Standard on Evaluation for the Government of Canada*;
- The structural elements required by the policy instruments—for example:
 - That the head of evaluation's position is classified appropriately;
 - That the head of evaluation has unencumbered access to the deputy head; and
 - That a departmental evaluation committee is in place;
- The content requirement, specifically the requirement for all five core issues to be addressed in every evaluation;
- The coverage requirement for evaluation of all direct program spending; and
- The frequency requirement for evaluation of all direct program spending every five years.

The external influences on the outputs are:

- The findings of the Centre of Excellence for Evaluation and the Office of the Auditor General related to evaluation quality, coverage and utility of evaluations; and
- The departments' experiences related to evaluation quality and utility.

In moving along the chain from outputs to immediate outcomes, the causal links are:

- The assumption that there is sufficient capacity (human resources, competence and financial resources) to meet the requirements;
- The assumption that there is an ability to attract, retain and train personnel; and
- The risk that the deputy head will require that other information needs be addressed.

The immediate outcomes of the 2009 *Policy on Evaluation* are:

- A robust, neutral evaluation function;
- All core issues are addressed in all evaluations;
- Comprehensive coverage;
- Five-year frequency of evaluations; and
- Departmental reports on the state of performance measurement.

The external influences on the immediate outcomes are:

- Performance measurement;
- The initial state of evaluation in the organization and pre-existing trends;
- The availability of competent evaluators and the ability to hire them;
- Fiscal constraints;
- A heightened need for evaluation information;
- Instability in the evaluation universe; and
- Financial information systems.

In moving along the chain from immediate outcomes to intermediate outcomes, the causal links are:

- The assumption of the development of evaluation-receptive attitudes and culture;
- The assumption that evaluation units are able to seize opportunities afforded by the policy;
- The risk that special demands from deputy heads sidetrack evaluations;
- The risk that limited capacity does not improve evaluation quality; and

- The risk that there is no internal client for an evaluation.

There are two sets of intermediate outcomes of the 2009 *Policy on Evaluation*.

The first set of intermediate outcomes is:

- Evaluations produce regular, credible and neutral information on the ongoing relevance and performance of direct program spending;
- Evaluations answer questions of interest to management;
- Evaluations produce findings about direct program spending entities that are useful for decisions;
- Evaluations produce findings when users need them; and
- Evaluations respond to emerging needs.

The external influences on the first set of intermediate outcomes are:

- Management Accountability Frameworks, including increased attention by the deputy head and senior managers, risk-averse evaluation behaviour, failure to use existing flexibilities, and policy application at cross-purposes with policy performance;
- The Program Alignment Architecture;
- The *Policy on Management, Resources and Results Structures*; and
- *Financial Administration Act* requirements.

In moving along the chain from the first set of intermediate outcomes to the second set of intermediate outcomes, the causal links are:

- The assumption that there is a commitment to improving through evaluation;
- The assumption that managers are willing to deal with negative results publicly;
- The assumption of the timely delivery of evaluations;
- The assumption that programs remain stable;
- The assumption that evaluation recommendations are actionable;
- The assumption that the communication of evaluation findings maximizes knowledge transfer;
- The assumption that support systems are in place to put evaluation findings into use;
- The risk that there is limited deputy head interest;
- The risk that there is limited space for program managers;
- The risk that evaluations go unused; and
- The risk that evaluation is marginalized by coverage-driven evaluations.

The second set of intermediate outcomes of the 2009 *Policy on Evaluation* is:

- Evaluations meet the needs of deputy heads and other users;
- Evaluations are used in expenditure management;
- Evaluations are used in Cabinet decision making;
- Evaluations are used in policy and program development and improvement;
- Evaluations are used for accountability and public reporting;
- The evaluation function is part of executive discussion and decision making; and
- The evaluation function produces useful information across programs and sectors.

The external influences on the second set of intermediate outcomes are:

- Strategic reviews;
- The trend to evidence-based decision making; and
- The evidence-mindedness of senior managers and parliamentarians.

In moving along the chain from the second set of intermediate outcomes to the ultimate outcome, the causal links are:

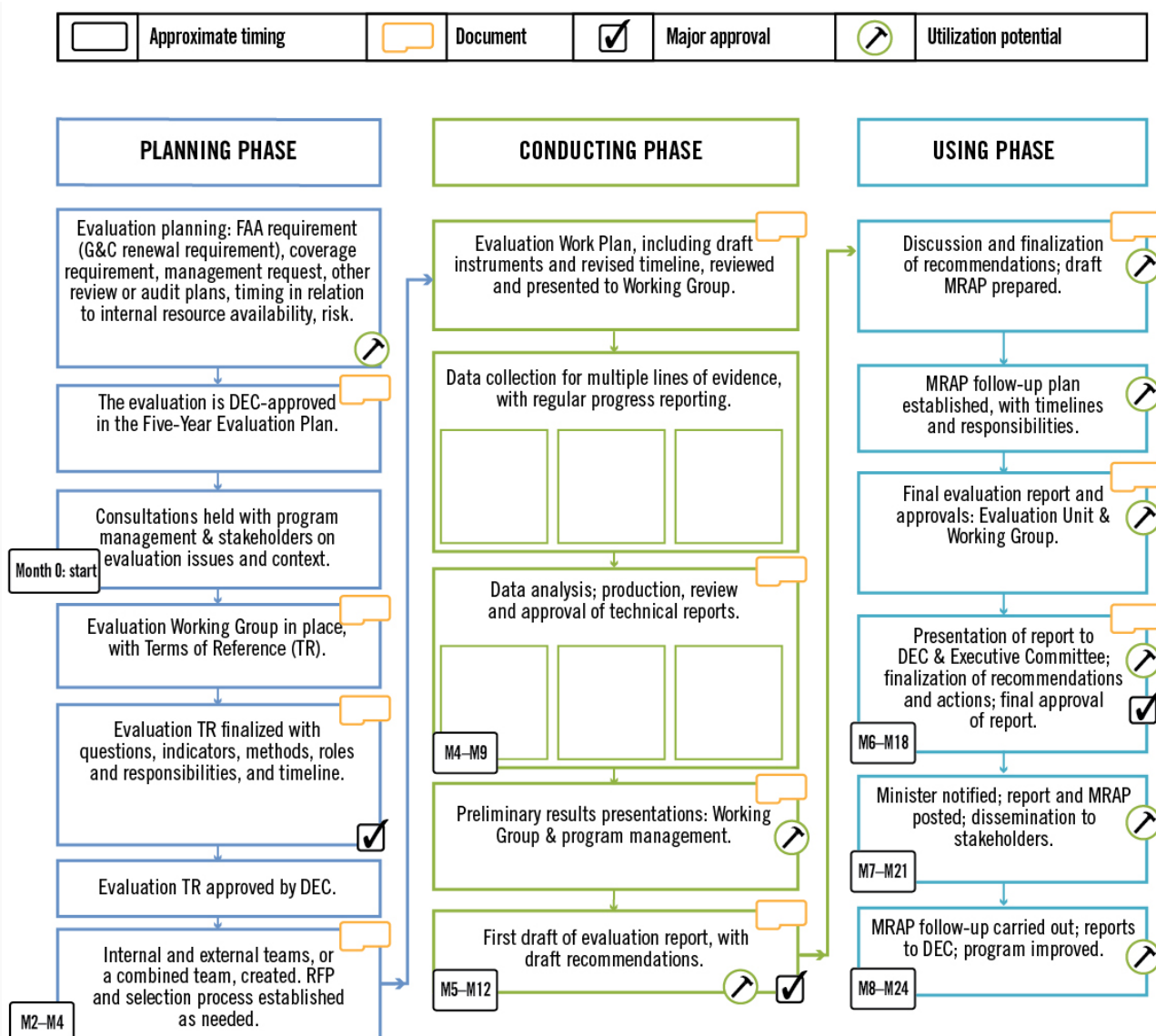
- The assumption that evaluation information is relevant and responsive to government management and accountability;
- The risk that evaluation information is disregarded in favour of other information or other decision-making influences; and
- The risk that the evaluation function is devalued.

The ultimate outcome of the 2009 *Policy on Evaluation* is that government is well managed and accountable and that resources are allocated to achieve results.

The external influences on the ultimate outcome are:

- Other forms of review and audit; and
- The renewal of the Expenditure Management System, which includes managing for results, disciplined proposals and strategic review.

Figure 8. Generic Evaluation Process Map of the Life Cycle of a Departmental Evaluation



▼ Figure 8 - Text version

The figure shows the activities or steps in the life cycle of a departmental evaluation. It also identifies the approximate timing, in months, for individual steps or groups of activities during an evaluation's life cycle; the activities or steps that result in the production of a document; the major approval steps associated with certain documents; and the steps in the life cycle where there are opportunities for evaluations or their preliminary products to be used.

The life cycle of a departmental evaluation has three phases: planning, conducting and using the evaluation.

The planning phase has seven steps:

Step one involves the planning of the evaluation. In this step, requirements such as the *Financial Administration Act* requirement (the grants and contributions renewal requirement) or a coverage requirement are identified. Consideration is given to any management requests, to other plans for review or audit, to the timing of the evaluation in relation to internal resource availability, and to risk.

Step two involves the listing of the evaluation in the five-year evaluation plan. Step two results in the production of the five-year evaluation plan, which is approved by the departmental evaluation committee.

Step three involves consultations with program management and stakeholders on the evaluation issues and context. In terms of evaluation timing, this step marks the start of the conduct of the evaluation, or month zero.

Step four involves the establishment of the evaluation working group and its terms of reference.

Step five, a major approval step, involves finalizing the terms of reference for the evaluation. The terms of reference document contains the evaluation questions, indicators, methodology, roles and responsibilities, and the timeline for the evaluation.

Step six involves the approval of the terms of reference for the evaluation by the departmental evaluation committee.

Step seven, the final step of the planning phase, involves the creation of internal and external evaluation teams, or a combined evaluation team. In addition, a request for proposal (RFP) and a selection process are established to select external team members, when required. The RFP is a document produced during this step. Step seven typically occurs between months two and four of the evaluation's life cycle.

The conducting phase has five steps.

Step one involves the review and presentation of the evaluation work plan, including the draft data collection instruments and revised timeline, to the working group. The evaluation work plan is a document produced during this step.

Step two involves the collection of data for multiple lines of evidence, with regular progress reporting.

Step three involves the analysis of data and the production, review and approval of technical reports. Technical reports are documents produced during this step, which typically occurs between months 4 and 9 of the evaluation life cycle.

Step four involves the presentation of preliminary results to the working group and to program management. The presentations are documents produced during this step.

Step five, the final step of the conducting phase, involves the production of the first draft of the evaluation report, including the draft recommendations. The draft report is a document produced during this step, which typically occurs between months 5 and 12 of the evaluation life cycle, and approved before moving into the using phase of the evaluation life cycle.

The using phase has six steps.

Step one involves discussion and finalization of the evaluation recommendations and the preparation, by program management, of a management response and action plan (MRAP). The MRAP is a document that is produced during this step.

Step two involves the establishment of a follow-up plan to the MRAP, including the timelines and responsibilities for actions.

Step three involves the production of the final evaluation report and its approval by the evaluation unit and the working group. The final evaluation report is a document produced during this step.

Step four involves the presentation of the report to the departmental evaluation committee and the executive committee, the finalization of recommendations and actions, and the final approval of the

evaluation report. This step typically occurs between months 6 and 18 of the evaluation life cycle, and represents a major approval step.

Step five involves the notification of the Minister regarding the evaluation, the posting of the report and its MRAP on the departmental website, and the dissemination of the report to stakeholders. This step typically occurs between months 7 and 21 of the evaluation life cycle.

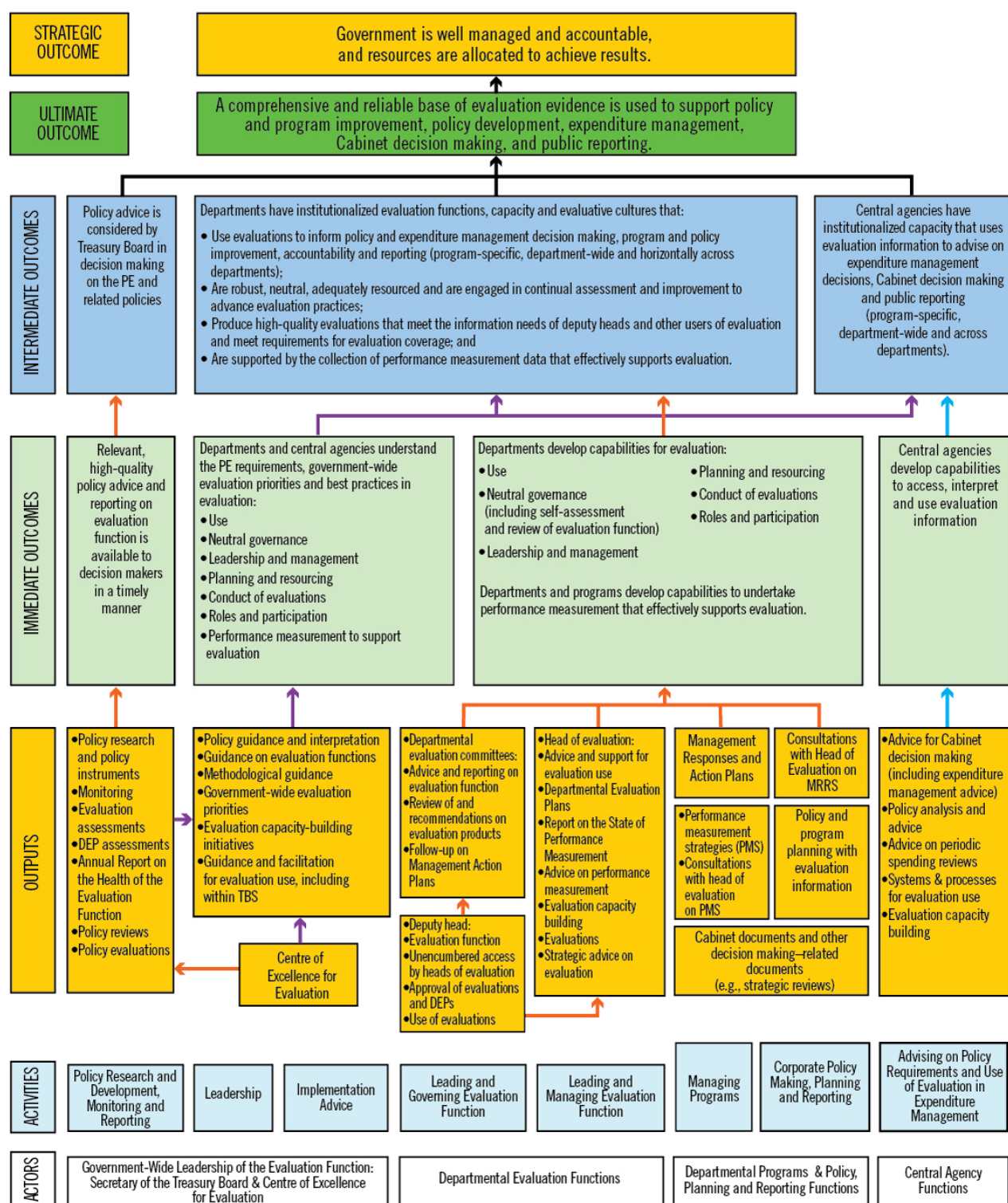
Step six, the final step of the evaluation life cycle, involves follow-up of the MRAP to ensure it is carried out, reporting to the departmental evaluation committee, and improvement of the program.

The process map identifies the opportunities for using evaluations during the evaluation life cycle. These opportunities occur during the following steps: evaluation planning (process use); presentation of preliminary results; draft reporting; discussion and finalization of recommendations and preparation of the MRAP; establishment of the follow-up plan to the MRAP; final evaluation report and approvals; presentation of report to the departmental evaluation committee and the executive committee, finalization of recommendations and actions; dissemination to stakeholders; MRAP follow-up and reporting to the departmental evaluation committee.

The generic process map was developed using information from the performance case studies.

The evaluation life cycle consists of three phases: planning, conducting and using. The case studies found that although the entire cycle can take up to two years, activities can be compressed if there is a need to complete an evaluation quickly. The longest period appears to be the approval process, between preliminary finding and final posting. Departments in the case studies generally update their evaluation plans annually, which may affect the timing and interplay between the evaluation phases. The Departmental Evaluation Committee (DEC) is engaged at a minimum of two points: the approval of the Departmental Evaluation Plan, and the approval of the final evaluation report. In addition, some departments engage their DEC to discuss the scope of evaluations and preliminary findings. The case studies identified numerous opportunities for using evaluations throughout the evaluation cycle. As shown in the process map, these opportunities occur during evaluation planning (process use), preliminary results presentations, draft reporting, discussion and finalization of recommendations, preparation of the Management Response and Action Plan (MRAP), final evaluation report and approvals, presentation or report to the DEC and Executive Committee; finalization of recommendations and actions, dissemination to stakeholders, MRAP follow-up and reports to the DEC.

Figure 9. Logic Model for Implementing the 2009 *Policy on Evaluation*



▼ Figure 9 - Text version

The figure shows the logic model for implementing the 2009 *Policy on Evaluation*, including the actors involved in implementation, the activities involved, the outputs of implementing the policy, and the immediate, intermediate, ultimate and strategic outcomes of the policy. The logic model also shows interactions among the elements.

The actors involved in policy implementation are the Secretary of the Treasury Board and the Centre of Excellence for Evaluation, which provide government-wide leadership to the evaluation function; departmental evaluation functions; departmental programs and policy, planning and reporting functions; and central agency functions.

The activities of the Secretary of the Treasury Board, supported by the Centre of Excellence for Evaluation, include policy research and development, monitoring and reporting; leadership; and implementation advice. In turn, the Centre of Excellence for Evaluation produces the following outputs: policy research and policy instruments, monitoring, evaluation assessments, assessments of departmental evaluation plans, Annual Reports on the Health of the Evaluation Function, policy reviews and policy evaluations; and policy guidance and interpretation, guidance on evaluation functions, methodological guidance, government-wide evaluation priorities, evaluation capacity-building initiatives, guidance and facilitation for evaluation use, including within the Treasury Board of Canada Secretariat.

The first immediate outcome associated with these outputs is that relevant, high-quality policy advice and reporting on the evaluation function is available to decision makers in a timely manner. The second immediate outcome is that departments and central agencies understand the requirements of the *Policy on Evaluation*, government-wide evaluation priorities and best practices in evaluation regarding use, neutral governance, leadership and management, planning and resourcing, conduct of evaluations, roles and participation, and performance measurement to support evaluation.

The intermediate outcome flowing from the first immediate outcome, “relevant, high-quality policy advice and reporting on the evaluation function is available to decision makers in a timely manner,” is that the policy advice is considered by the Treasury Board in decision making about the *Policy on Evaluation* and related policies.

The second immediate outcome contributes to two intermediate outcomes. The first intermediate outcome is that departments have institutionalized evaluation functions, capacity and evaluative cultures that use evaluations to inform policy and expenditure management decision making, program and policy improvement, accountability and reporting (program-specific, department-wide and horizontally across departments); are robust, neutral, adequately resourced and are engaged in continual assessment and improvement to advance evaluation practices; produce high-quality evaluations that meet the information needs of deputy heads and other users of evaluation and meet requirements for evaluation coverage; and are supported by the collection of performance measurement data that effectively supports evaluation.” The second intermediate outcome is that central agencies have institutionalized capacity that uses evaluation information to advise on expenditure management decisions, Cabinet decision making and public reporting (program-specific, department-wide and across departments).

Departmental evaluation functions have two types of activities: leading and governing the evaluation function, and leading and managing the evaluation function. The three groups of outputs produced by departmental evaluation functions are: (1) deputy heads establish evaluation functions, ensure unencumbered access by heads of evaluation to deputy heads, approve evaluations and departmental evaluation plans, and ensure use of evaluations; (2) departmental evaluation committees give advice and report on the evaluation function, review and recommend evaluation products for approval by the deputy head, and follow-up on Management Action Plans; and (3) heads of evaluation provide strategic advice on evaluation, give advice and support for evaluation use, produce departmental evaluation plans and Reports on the State of Performance Measurement, advise on performance measurement, provide support to evaluation capacity building, and carry out evaluations.

Departmental programs and policy, planning and reporting functions have two types of activities: managing programs; and corporate policy making, planning and reporting. These activities produce the following five groups of outputs: (1) Cabinet documents and other decision making–related documents (for example, strategic reviews); (2) performance measurement strategies and consultations with heads of evaluation on performance measurement strategies; (3) policy and program planning with evaluation information; (4) management responses and action plans; and (5) consultations with heads of evaluation on Management, Resources and Results Structures.

The activities and outputs of departmental programs and policy, planning and reporting functions, in combination with the activities and outputs from departmental evaluation functions, lead to the following set of immediate outcomes: (1) departments develop capabilities for evaluation, including capabilities related to use, neutral governance (including self-assessment and review of evaluation function), leadership and management, planning and resourcing, conduct of evaluations, and roles and participation; and (2) departments and programs develop capabilities to undertake performance measurement that effectively supports evaluation.

The immediate outcomes contribute to one of the intermediate outcomes aligned with the activities and outputs of the Secretary of the Treasury Board and the Centre of Excellence for Evaluation, which was previously described as “departments have institutionalized evaluation functions, capacity and evaluative cultures that use evaluations to inform policy and expenditure management decision making, program and policy improvement, accountability and reporting (program-specific, department-wide and horizontally across departments); are robust, neutral, adequately resourced and are engaged in continual assessment and improvement to advance evaluation practices; produce high-quality evaluations that meet the information needs of deputy heads and other users of evaluation and meet requirements for evaluation coverage; and are supported by the collection of performance measurement data that effectively supports evaluation.”

Central agency functions have one activity, which consists of advising on policy requirements and the use of evaluation in expenditure management. This activity leads to the following set of outputs: advice for Cabinet decision making (including expenditure management advice); policy analysis and advice; advice on periodic spending reviews; systems and processes for evaluation use; and evaluation capacity building. In turn, this set of outputs lead to one immediate outcome: central agencies develop capabilities to access, interpret and use evaluation information. This immediate outcome contributes to one of the intermediate outcomes aligned with the activities and outputs of the Secretary of the Treasury Board and the Centre of Excellence for Evaluation, which was previously described as “central agencies have institutionalized capacity that uses evaluation information to advise on expenditure management decisions, Cabinet decision making and public reporting (program-specific, department-wide and across departments).”

The three intermediate outcomes described combine to produce one ultimate outcome: “A comprehensive and reliable base of evaluation evidence is used to support policy and program improvement, policy development, expenditure management, Cabinet decision making, and public reporting.” This ultimate outcome leads to one strategic outcome: “Government is well managed and accountable, and resources are allocated to achieve results.”

Endnotes

- 1 Scope refers to the extent of spending that is included in an evaluation; the scope may be limited to an individual program or may group together programs or spending.
- 2 Focus refers to the issues addressed in an evaluation.
- 3 The five core issues, as listed in [Annex A of the Directive on the Evaluation Function](#), are as follows: continued need for the program, alignment with government priorities, alignment with federal roles and responsibilities, achievement of expected outcomes, and demonstration of efficiency and economy.
- 4 [2006 November Report of the Auditor General of Canada, Chapter 1: "Expenditure Management System at the Government Centre."](#)
- 5 Report of the Standing Committee on Public Accounts, [The Expenditure Management System at the Government Centre and the Expenditure Management System in Departments](#), February 2008.
- 6 The Standing Committee on Public Accounts held meetings in February 2007 to collect evidence for its report; the completed report was adopted in February 2008.
- 7 The 2009 *Policy on Evaluation* defines direct program spending as the portion of total budgetary spending that excludes public debt charges and major transfers to persons and other levels of government. Direct program spending includes operating and capital spending and grants and contributions, as specified in the Public Accounts of Canada.
- 8 [The Magenta Book: Guidance for Evaluation](#), HM Treasury, April 2011.
- 9 ["Increased Emphasis on Program Evaluations,"](#) Executive Office of the President, Office of Management and Budget, M-10-01, Memorandum for the Heads of Executive Departments and Agencies, October 9, 2009 (PDF Version, 89 KB).
- 10 Coverage requirements in the 2009 policy pertain to covering direct program spending rather than programs. This requirement provides more implementation flexibility to departments, as programs do not necessarily need to be evaluated one at a time. Direct program spending refers to the portion of total budgetary spending that excludes public debt charges and major transfers to persons and other levels of government. Direct program spending includes operating and capital spending and grants and contributions, as specified in the Public Accounts of Canada.
- 11 The former Treasury Board Portfolio Advisory Committee (TBPAC) of deputy ministers was consulted in October 2008 about the draft 2009 *Policy on Evaluation*. This committee has been renamed the Public Service Management Advisory Committee (PSMAC).
- 12 In calibrating the level of evaluation effort, departments take into account their information needs, program characteristics and risks, as well as the quality of performance information already available for the program.
- 13 In retrospective pretest design, information (for example, perceptions) that relates to before and after the program is collected at the same time.

- [14](#) See the definition in [Appendix A of the Directive on Transfer Payments](#).
- [15](#) See the definition in [Appendix A of the Directive on Transfer Payments](#).
- [16](#) See [Appendix H of the Directive on Transfer Payments](#).
- [17](#) Respondents answered on a four-point scale: 1 – very useful, 2 – somewhat useful, 3 – not very useful, or 4 – not at all useful.
- [18](#) In a sample of departmental evaluation plans submitted to the Centre of Excellence for Evaluation in 2011, two thirds of evaluations had scopes that were aligned with the level of a Program Alignment Architecture Program or Sub-Program.
- [19](#) Evidence from the Implementation Review showed that a number of departments reported that they had purposely minimized their support for the development of performance measurement strategies in order to focus limited resources on meeting coverage requirements. In contrast, a smaller number of departments reported that they had purposely expanded their support for performance measurement.
- [20](#) Refers to evidence from the Implementation Review.
- [21](#) Refers to evidence from performance case studies, stakeholder consultations, the survey of evaluation managers and evaluators, the Implementation Review and the Secretariat's annual Capacity Assessment Surveys of departmental evaluation functions.
- [22](#) In the 2008–09 Capacity Assessment Survey, the highest response category available to surveyed departments was “almost always”; there was no distinct response category for “always.” Any department that always considered evaluations in their Treasury Board submissions and Memoranda to Cabinet responded “almost always.” For comparative purposes, the reported percentages for responses to this question in the 2012–13 Capacity Assessment Survey include those that responded “all” or “almost all.”
- [23](#) Both the performance case studies and the examples of evaluation use that departments described when responding to the 2012–13 Capacity Assessment Survey showed widespread use to support renewals of existing spending—notably, for ongoing programs of grants and contributions.
- [24](#) The Implementation Review documented medium to high use of evaluations in spending review processes from 2007–08 to 2010–11, with usage peaking in 2010–11 (91% of reporting departments).
- [25](#) Evaluators reported that the 2009 policy had a relatively high impact on increasing the overall utility of evaluations for spending reviews, primarily because of the requirements for comprehensive coverage and core issues.
- [26](#) Refers to Secretariat program sector analysts consulted during the Implementation Review.
- [27](#) Refers to the findings of the 2003 interim evaluation of the 2001 *Evaluation Policy*.

- [28](#) Refers to the findings of *Deputy Head Consultations on the Evaluation Function*, Centre of Excellence for Evaluation, January 2011.
- [29](#) According to 84 of 112 reports that the Centre of Excellence for Evaluation assessed for the Management Accountability Framework Assessment process in 2010–11.
- [30](#) For example, one evaluation that identified a program's challenges in accessing accurate project information led to the launch of an organization-wide project management database. Another evaluation that provided information about the “reach” of one program to its target groups led to changes in the targeting of a new program.
- [31](#) Almost two thirds of program managers surveyed (64%) reported that evaluations conducted since 2009 were somewhat useful or very useful for public reporting. Program managers surveyed, and even more so, evaluators surveyed, attributed a positive impact to the policy and each of its three requirements (comprehensive coverage of direct program spending, five-year frequency for evaluations, and examination of the five core issues) on the utility of evaluations to support accountability and public reporting.
- [32](#) See Appendix D for an illustration of the key conditions, as shown by the contribution theory of the 2009 *Policy on Evaluation*.
- [33](#) Refers to heads of evaluation, departmental evaluation committee members and program managers consulted for the Implementation Review.
- [34](#) Refers to evidence from the Implementation Review.
- [35](#) An additional 51% of program managers surveyed reported that departmental dialogue on evaluation had remained the same.
- [36](#) The 2013–14 Capacity Assessment Survey showed that 97% of large organizations reported internally on the implementation of action plan items at specified times during the fiscal year.
- [37](#) The performance case studies and the Implementation Review found that senior management paid increased attention to the implementation of action plans. Further, the majority of evaluators surveyed reported that support from senior management and the engagement of departmental evaluation committees had positive impacts on evaluation use.
- [38](#) Refers to evidence from the survey of evaluators. Using their experiences since 2009, evaluators rated the impact of a series of factors on the use and utility of evaluations.
- [39](#) Refers to evidence from the survey of evaluators, which showed that 81% of evaluators reported that calibrating the evaluation scope and approach according to program risks was somewhat effective or very effective for expanding evaluation coverage; 80% of evaluators reported that aligning the scope of evaluations with Program Alignment Architecture units was somewhat effective or very effective; 74% of evaluators reported that clustering related programs was somewhat effective or very effective; and 81% of evaluators reported that increasing the use of internal staff to conduct evaluations was somewhat effective or very effective.

- [40](#) The [2012 Annual Report on the Health of the Evaluation Function](#) reported that government-wide, departments decreased their spending on external consultants.
- [41](#) Refers to stakeholder consultations.
- [42](#) Refers to evidence from the survey of evaluators.
- [43](#) Refers to stakeholder consultations, including with deputy heads, and to case studies.
- [44](#) Refers to consultations with heads of evaluation and case study consultations with central agency representatives and program managers, the [2013 Spring Report of the Auditor General of Canada, Chapter 1: “Status Report on Evaluating the Effectiveness of Programs,”](#) and the Implementation Review.
- [45](#) Evidence from the document review showed that the trend toward evaluating larger entities predated the 2009 policy. The [2003 interim evaluation of the 2001 Treasury Board Evaluation Policy](#) found that during the previous two years, in a sample of large organizations, 27% of evaluations covered broader policies and initiatives. However, it is unknown whether these entities were comparable to Program Alignment Architecture units.
- [46](#) The legal requirement had to be met for the first time by the end of December 2011, whereas the policy's requirement for comprehensive coverage of direct program spending came into effect in April 2013, and did not need to be fully met until March 2018. The majority of examples that departments provided for the case studies were ongoing programs of grants and contributions, for which evaluation requirements are set by the [Financial Administration Act](#), the [Policy on Transfer Payments](#) and the [Policy on Evaluation](#).
- [47](#) The six countries were Australia, the United Kingdom, the United States, India, Mexico and Switzerland.
- [48](#) This finding was consistent with evidence from the Implementation Review.
- [49](#) Refers to the evidence from the surveys of program managers and evaluators and from case studies.
- [50](#) Evaluators who were surveyed on the impacts of the comprehensive coverage requirement reported that the requirement positively impacted the following uses, in order from most often cited to least often cited: accountability, spending reviews, program or policy improvement, public reporting, Cabinet decision making, program or policy development, new resource allocations, renewal of existing resource allocations, and internal resource allocation.
- [51](#) Refers to consultations conducted with deputy heads in 2014.
- [52](#) Refers to case study consultations and the survey of program managers.
- [53](#) Refers to case study consultations and the survey of evaluators.
- [54](#) These variations were noted during the Implementation Review.

- [55](#) Central agencies' higher level of support for the five-year requirement may reflect the transactional nature of their work (for example, analyzing program renewal requests, often on a five-year basis) and the specific support that evaluations can provide to them. Central agencies noted that unless spending decision cycles and evaluation cycles were aligned, evaluations could sometimes be rendered moot. When evaluations were considered to be out of date, central agency analysts would seek more current information from other sources.
- [56](#) The length of time it takes to complete evaluations affects how easily they fit into a five-year cycle. Before the 2009 policy, a diagnostic study ("Evaluation Policy Renewal: Diagnostic," Treasury Board of Canada Secretariat, PowerPoint presentation, June 28, 2006) found that 47% of evaluations took more than a year to complete. Stakeholder consultations and case studies of evaluations of grants and contributions programs showed that lengthy or complex evaluations could mean that one evaluation of the program had hardly finished before the next one had to begin, creating the perception that the subsequent evaluation was premature. Delays in contracting, fieldwork and approvals sometimes contributed to lengthy evaluations. Although most user groups saw evaluation units making efforts to be timely, sometimes evaluations were not available when key decisions were being made.
- [57](#) [2013 Spring Report of the Auditor General of Canada, Chapter 1: "Status Report on Evaluating the Effectiveness of Programs."](#)
- [58](#) Refers to case studies, deputy head consultations, head of evaluation consultations, and the survey of evaluators.
- [59](#) Refers to evidence from case study consultations.
- [60](#) Refers to evidence from case studies, consultations with heads of evaluation, and the survey of evaluators.
- [61](#) The Implementation Review found that the requirement for a consistent set of evaluation issues lowered the likelihood of information gaps. The standardization of issues was considered to be particularly important for supporting the work of Secretariat analysts and for the planning of horizontal evaluations.
- [62](#) Respondents answered on a four-point scale: 1 – very useful, 2 – somewhat useful, 3 – not very useful, or 4 – not at all useful.
- [63](#) Refers to evidence from case studies, consultations with heads of evaluation, and the survey of evaluators.
- [64](#) This is consistent with the findings of the Implementation Review, which identified that having to cover all five core issues in evaluations can constrain the ability of departments to meet the information needs of deputy heads, which are sometimes outside the core issues.
- [65](#) For example, a similar concern was raised by the [2013 Spring Report of the Auditor General of Canada, Chapter 1: "Status Report on Evaluating the Effectiveness of Programs."](#) The Auditor

General reported that among the audited departments, covering all five core issues is not always useful, depending on program managers' and decision-makers' needs.

- [66](#) These concerns were observed through the case studies and other data sources including the Implementation Review and the heads of evaluation consultation.
- [67](#) In case study consultations and in consultations with heads of evaluation, the core relevance issues were most often identified as not applicable in specific evaluations. Core issue 3 (alignment with federal roles and responsibilities) and, to a lesser extent, core issue 2 (alignment with federal government priorities) were the ones most frequently identified as difficult to address or of limited utility. This finding was consistent with the results of the surveys of evaluators and program managers. However, central agency users were more likely to find these issues useful.
- [68](#) In case studies, these issues were less applicable in situations where relevance was unquestioned—for example, where the program was new or stable, where the deputy had no interest in examining relevance, or where relevance was already clearly established by legislation, by a recent review or by another means.
- [69](#) Refers to evidence from the Implementation Review, case studies and consultations, including those with deputy heads.
- [70](#) Under the policy, small departments are required to address the core issues only when they conduct evaluations of ongoing grants and contributions programs.
- [71](#) Refers to evidence from stakeholder consultations and the Implementation Review.
- [72](#) This was documented by the document review.
- [73](#) Other evaluation issues included governance and alternative delivery models, adequacy of performance measurement, strategic and policy issues, unintended outcomes, lessons learned and success factors, cost-effectiveness, and progress since the last evaluation.
- [74](#) Refers to evidence from case studies and stakeholder consultations.
- [75](#) Surveyed program managers and evaluators indicated that about a quarter of non-core issues could be subsumed by the existing core issues.
- [76](#) The case studies provided additional insight, showing that although the Secretariat communicated the flexibilities around addressing core issues, departments worried that Management Accountability Framework assessment methods and ratings would not reflect favourably on departments that used the flexibilities. Because Management Accountability Framework assessment ratings have an impact on the performance assessments of deputy heads and have repercussions for departmental management teams, the prospect of low ratings was a disincentive for using flexibilities related to the *Policy on Evaluation*.
- [77](#) For example, they calibrated by restricting data sources to a high-level document review, excluding these questions when collecting data from stakeholders, and not benchmarking to, or comparing with, similar programs.

- [78](#) 1996 May Report of the Auditor General of Canada, Chapter 3: “Evaluation in the Federal Government.”
- [79](#) Refers to evidence from the Implementation Review and analysis of case studies.
- [80](#) The 2013–14 Capacity Assessment Survey showed that 97% of large organizations reported internally on the implementation of action plan items at specified times during the year, and that 65% used a systematic process for reporting the impacts of evaluation to their deputy heads. Case study and stakeholder consultations also demonstrated that departments had systematic processes in place to report to departmental evaluation committees and deputy heads on the follow-up of action plans.
- [81](#) Other lines of evidence included the Implementation Review, which found evidence of process use where corrective actions were implemented before the management response and action plan were finalized. Instances were also noted where evaluation information that was not part of formal recommendations was used by program managers or senior management, sometimes in very significant ways, to make improvements or to inform other decisions.
- [82](#) Refers to evidence from the surveys of program managers and evaluators.
- [83](#) These were not departments that were included in consultations or case studies.
- [84](#) Refers to evidence from case study consultations.
- [85](#) Treasury Board policy circular on “Evaluation of Programs by Departments and Agencies,” 1977.
- [86](#) “Evaluation and Audit,” in the *Treasury Board Manual*, 1992.
- [87](#) “Review, Internal Audit and Evaluation,” in the *Treasury Board Manual*, 1994.
- [88](#) [Study of the Evaluation Function in the Federal Government](#), Centre of Excellence for Evaluation, April 2004.
- [89](#) [Assessing Program Resource Utilization When Evaluating Federal Programs](#), Treasury Board of Canada Secretariat, 2013.
- [90](#) [From Red Tape to Clear Results: The Report of the Independent Blue Ribbon Panel on Grants and Contributions Programs](#) (PDF Version, 593.1 KB).
- [91](#) Report of the Standing Committee on Public Accounts, [The Expenditure Management System at the Government Centre and the Expenditure Management System in Departments](#), February 2008.
- [92](#) 1993 Report of the Auditor General of Canada, Chapter 9: “Program Evaluation in Departments—The Operation of Program Evaluation Units”
- [93](#) 1996 May Report of the Auditor General of Canada, Chapter 3: “Evaluation in the Federal Government”

- [94](#) [2000 December Report of the Auditor General of Canada, Chapter 20: Managing Departments for Results and Managing Horizontal Issues for Results.](#)
- [95](#) [2009 Fall Report of the Auditor General of Canada, Chapter 1: "Evaluating the Effectiveness of Programs."](#)
- [96](#) [2013 Spring Report of the Auditor General of Canada, Chapter 1: "Status Report on Evaluating the Effectiveness of Programs,"](#)
- [97](#) Following completion of the Implementation Review and corrections to data that were self-reported by departments, the Secretariat later reported a corrected figure of 477 full-time equivalents in 2011–12.
- [98](#) See endnote 97.
- [99](#) [Section 2.2 of the 2012 Annual Report on the Health of the Evaluation Function](#), Treasury Board of Canada Secretariat.
- [100](#) The 2009–10 data were based on a sample of 21 evaluations that were assessed by the Centre of Excellence for Evaluation regarding the sufficiency of performance measurement data. The 2011–12 data were based on a sample of 101 evaluations.

► [Report a problem or mistake on this page](#)

[!\[\]\(2bdfe261b986065ee0ac76460d6528c9_img.jpg\) Share this page](#)

Date modified: 2015-05-13

[Contact us](#)

[Departments and agencies](#)

[Public service and military](#)

[News](#)

[Treaties, laws and regulations](#)

[Government-wide reporting](#)

[Prime Minister](#)

[How government works](#)

[Open government](#)

- [Social media](#)
- [Mobile applications](#)
- [About Canada.ca](#)

- [Terms and conditions](#)
- [Privacy](#)

[Top of page](#) 

Canada 