

Program Evaluation and Aboriginal Affairs: A History and a Thought Experiment*

Ian D. Clark and Harry Swain

The evaluation function at the Department of Indian Affairs and Northern Development (DIAND)¹ is now working as well as any in the Canadian federal government, fully up to the demands of the central agencies. These standards are a pale but expensive shadow, however, of the ambitious age of High Modernism forty years ago. Rod Dobell's approach has matured since then, but in key respects he is still at odds with the degraded standards of the present system.

A THOUGHT EXPERIMENT

The present performance of the evaluation function at DIAND stems from the central agency pressure to embed evaluation in the policy and expenditure management system, which owes so much to Dobell and his colleagues at the Treasury Board 40 years ago. It also owes much to the complexity and confusions of policy toward Canada's aboriginal peoples.

Since leaving Ottawa in 1976 for Paris and then Victoria, Dobell has continued to think and write on a wide range of policy issues, including the philosophy and practice of evaluation and the relationship between Canadian governments and aboriginal Canadians. To frame this essay on the evolution of program evaluation in aboriginal affairs, we pose the question: *What if Rod Dobell had stayed in Ottawa?*

Suppose Dobell stayed on in the positions that we once held, Secretary of the Treasury Board and Deputy Minister of Indian Affairs and Northern Development respectively. Although we acknowledge that Rod is intelligent enough to have done both our jobs at the same time, this would have been structurally inappropriate. Our experiment has Rod rotating every five years

* NOTE: This is the penultimate draft of a chapter in an edited volume "A Subtle Balance: Expertise, Evidence, and Democracy in Policy and Governance, 1960-2010," which is being published in Spring 2015 by McGill-Queens University Press (MQP), and which has given the authors permission to post this draft.

between 1978 and 2008 from deputy at Indian Affairs to Treasury Board secretary. He could thus periodically oversee the government-wide policy on evaluation, influence the resources devoted to evaluation, and actually use evaluation resources in a real department. In 2013, after seven five-year rotations, Rod would at last have earned the right to retire to Victoria.

Our point is not to imply that any particular individuals have failed in their duty but to illustrate how we think the worlds of evaluation and aboriginal affairs might have been different if the federal government had had the benefit of the sustained contribution of senior officials who were as gifted in the theory, practice and exposition of policy analysis and evaluation² as Rod Dobell.

One of the realistic elements in our experiment is that these matters – evaluation policy, overall evaluation resources, and the uses to which evaluations are put – are almost wholly in the hands of the mandarinate. All ministers we have worked with took these matters to be questions of management practice and were comfortable leaving the details to their officials. If deputy ministers decide that evaluation is important it is relatively easy to cast the function in a way that fits the ideological mold of whichever party is in power. We would assert that the current strengths and weaknesses of the evaluation function in the federal government today are almost solely attributable to the decisions and actions of officials rather than politicians.

HIGH MODERNISM IN OTTAWA

Summarizing Zussman's detailed history above, the new Trudeau government established a Planning Branch in the Treasury Board Secretariat in 1969, where Al Johnson became Secretary. Johnson recruited Douglas Hartle from the political economy department at the University of Toronto to head the branch and Hartle recruited many others, Dobell senior among them, to

systematically apply the ideas of rational budgeting and evaluation of policy alternatives that had been developed in the 1950s and 1960s in the United States.³ The central thought was that the costs of government actions could be measured against the outcomes of those actions, allowing rational re-design, a choice of different means, or a ranking of different action-outcome sets. This optimistic belief underlies any number of specific programs with acronyms like PPBS, MBO and PEMS.

In retrospect, the Ottawa of Hartle and Dobell can be considered the apogee of what Moran, Rein, and Goodin (2006) refer to as the “high modernism” stage of policy analysis, with its “technocratic hubris, married to a sense of mission to make a better world” and “an overwhelming confidence in our ability to measure and monitor that world.” They conclude that “high modernism is an anachronism.” In a similar vein, Dobuzinskis, Howlett and Laycock (2007) describe the divergence that has emerged between the “positivist” approach to policy analysis – rooted in micro-economics and based on mastery of a number of formal analytical techniques with the intent of “speaking truth to power” – and a newer “post-positivist” approach focusing more on process-related techniques for affecting policy discourse, ideas and arguments. Van Loon summarizes the fate of the Treasury Board planning system centred in the Hartle/Dobell Planning Branch:

Reams of material ha[ve] been produced on the difficulties encountered by microeconomic and evaluation-based planning processes. Indeed, a significant quantity of that material has been produced by Hartle himself who quickly grew disenchanted with the potential for introducing the particular definition of rationality he had espoused into political decision-making unless the rules of the game and the incentives and behaviour

patterns of politicians and bureaucrats could be drastically altered. Suffice it to say here that the Planning Branch planning system never took hold in Ottawa, nor even managed to dominate its own agency, the Treasury Board. Thus, although some excellent technical work was done by its evaluators and some very interesting findings emerged, the high tide of the Planning Branch system ebbed quickly. The Planning Branch itself disappeared in an austerity move in 1978 and the evaluation function was shifted to the Office of the Comptroller General. (1981, 197)

Both Hartle and Dobell were brilliant, fearless and analytically rigorous. Their intellectual journey, portrayed in writings about what they learned from the Planning Branch's limited success in applying political economic logic to real government settings, has been insightfully described in Dobell's contribution to the book written in tribute to Hartle (Dobell 1999). He calls the 1968-1978 period the culture of BOISE – BOLD Inference and Social Engineering. This was the era of Cabinet Evaluation Studies, where major government expenditure programs whose economic outcomes were radically lower than stated expectations – such as the dairy program, investments in atomic energy, regional development incentives programs, and the Cape Breton Development Corporation – were examined with modern microeconomic techniques with the explicit object of trying to improve their effectiveness through changes in the program design. Dobell succeeded Hartle as head (Deputy Secretary) of the Planning Branch in 1973. Although the Branch was wound up soon after Dobell moved to the OECD, the underlying idea had the power of persistence. After a period of confusion in the late 1970s and early 1980s, the idea of program evaluation as a more modest and more routine part of governance began to be widely accepted. Dobell describes how BOISE evolved into the

more prosaic ROME – Results-Oriented Management and Evaluation – in the decade from 1988-1998 (Dobell 1999).

The notion of routine program evaluation became institutionalized in the expectations that central agencies (especially the Board) had for line departments, and in departmental structures. Large-scale expenditure reviews in 1985 and 1995 created a demand for evaluation studies and the loss-of-control scandals of the 1997-2007 period multiplied central rules. Crucial to the government's current central control system, the Management Accountability Framework (Treasury Board Secretariat 2012), is the idea of universal, periodic and ostensibly risk-based audit and evaluation, with attendant performance measurement apparatus.

RATIONALISM AND REJECTION IN ABORIGINAL AFFAIRS

The rationalism of the old Planning Branch was echoed in odd ways in the world of Indian Affairs. Until 1970, policy was unabashedly assimilationist. Under the *Indian Act* the federal government acted as the benevolent ward of indigenous peoples whose innocence of modern ways, it was thought, was so severe that they had to be protected from European settlers. Indian agents controlled every minute facet of community life on the reserves to which they had been consigned. Escape from these prisons of paternalism required assimilation and the casting aside of traditional languages and ways of life. In 1969 Prime Minister Trudeau and his young Indian Affairs minister, Jean Chrétien, appalled at the human consequences of this policy (Hawthorn 1966-67), published a White Paper which proposed to end at a stroke Canada's internal colonialism. Race-based discrimination (as well as entitlements) stemming from the obsolete and unequal treaties of the 19th century would end, and we would all be Canadians together. But

Indians objected. If the conditions of their lives were to be changed, they wanted a hand in the design. Harold Cardinal's Red Paper of 1970 put paid to Trudeau's Enlightenment rationalism.

The second great blow to assimilationist policy, quickly following Cardinal, came from the Supreme Court, which was unable to conclude in *Calder* (1973) that aboriginal rights had been extinguished in those parts of the country, like Chief Calder's Nisga'a territory, not covered by treaties. Suddenly Indians had important (and after 1982, constitutionally protected) land rights. And their circumstances were straining the country's conscience.

Federal policy leapt into an entirely new dimension. Expenditures increased greatly, though not as rapidly as the respect (and subventions) granted such organizations as the National Indian Brotherhood (later the Assembly of First Nations) and the Inuit Tapirisat. In 1981-82, pressure from these newly empowered groups and progressive Canadians generally resulted in a section of the new Charter of Rights and Freedoms recognizing existing treaty and aboriginal rights, whatever they were. For the next 15 years the Supreme Court began to flesh out the definition. Governments responded. By the late 1980s the negotiation of modern treaties no longer demanded the "extinguishment" of aboriginal rights; by 1995 the government had recognized as one of those rights an "inherent right of self-government." The separate but unequal system of fiscal apartheid, through which Canada encouraged indigenous peoples to stay in their traditional, often poorly endowed, territories through large payments to Indian Act, rather than traditional, governments, was cemented in place. And in parallel with the re-recognition of these old rights the programming of the Department of Indian Affairs moved away from direct delivery – initially to devolution to band governments under annual, later multi-year, contribution agreements, and then to transfer payments to new self-governing entities. Entirely aside from the pressure from the center, DIAND developed an intense interest in knowing

whether these new instrumentalities were effective. (Efficiency and economy claimed some rhetorical space but were distinctly secondary.)

EVALUATION AT INDIAN AFFAIRS AND NORTHERN DEVELOPMENT

By 1982 DIAND's evaluation practice was well established. The department had in fact anticipated TBS' 1977 direction by establishing the function as part of its own corporate policy in 1976. In keeping with the statement that the deputy minister was the primary client for program evaluations, only activities big enough (in several senses) to warrant that grandee's attention were to be covered. And since the deputy minister and, through him, the policy and expenditure management committees supporting the Cabinet were the primary audiences, there was an emphasis on identifying factors relevant to policy change and resource allocation priorities. The director of the evaluation branch was invited to append recommended changes to policy or program design to the "objective" studies going to the deputy. The benefits and costs of alternative program delivery systems, as well as the continuing relevance of program objectives, were seen as "basic issues for evaluation," to be covered in every study. A forward-looking, policy development outlook flavoured the whole enterprise, reflecting the optimism of the Dobell Planning Branch. TBS and the Comptroller General also received copies of completed work, and the Board might demand evaluations relevant to TB submissions. The central agencies were copied on the rolling five-year plans for evaluation studies so they could be taken into account in the Policy and Expenditure Management System, and Multi-Year Operational Plans. Central evaluations might have disappeared, but evaluation directorates in the line departments were given substantial scope, including an invitation to recommend policy changes, and generally encouraged to fan the flame of high modernism. The demise of the ministries of state during the short premiership of John Turner slackened demand for internal program evaluation.

The Mulroney government, elected in 1984, determined to reduce the Trudeau administration's bequest of alarming deficits and debt, undertook a system-wide expenditure review under deputy prime minister Erik Nielsen.⁴ The volume on aboriginal programs was entrusted mostly to strong-minded non-governmental theorists who had forgotten the lessons of 1970, however, and their recommendations gained no traction.

By the late 1980s the audit and evaluation functions in DIAND were overseen by a department-wide committee chaired by the associate deputy minister, and a steady stream of reports ensued, focussing on areas that were most problematic or were thought to be subject to imminent policy review. Methodologically the reports were not state-of-the-art, and performance indicators, while discussed, were not systematically developed. Senior management was all too aware of the fragility of the statistical base on which most performance measures rested. The spirit was willing, in other words, but the consequences of evaluations were modest.

It was not until a new government, under Prime Minister Jean Chrétien, pressed to the wall by debt and deficit, undertook an emergency program review in 1995, that much effect was felt on DIAND budgets.⁵ Minor short-term cuts and personnel losses were the focus of much attention, but the real impact on Indian people came through the setting of a capped rate of increase of two percent for all the department's social programs, most crucially education. Two decades years later the cap is still mostly in place, despite inflation and rapid population growth rates. Evaluation studies have gingerly pointed out the consequences of this sweeping budgetary decision, but to little avail. The lesson seems to be that the hammer of fiscal necessity overwhelms the pinpricks of program evaluations.

In 2006, a new deputy minister and a new chief of audit and evaluation began to revamp and modernize the function in the department. With the following wind of policy statements and assistance from TBS, and by its edict that all programs had to be evaluated if their funding was to continue, an expanded and methodologically better founded evaluation system was installed. Key features include (a) risk and materiality based planning of evaluations, (b) the timing of evaluations to fit the policy planning cycle, (c) base rather than soft funding of the evaluation unit, and (d) oversight by an Evaluation, Performance Management and Review Committee chaired by the deputy minister and salted with knowledgeable external members.⁶ The principal area for further work (now underway) is in the development of a sparse but robust system of performance measures covering all programs and the department as a whole – a task complicated by the disappearance of Statistics Canada’s mandatory long-form census.

These admirable features are weakened, however, by directions from the centre on how evaluations should be performed and presented. Every study must conclude whether the program is still relevant to the government’s goals (we cannot remember ever seeing a negative response); must examine program modalities but never question “policy;” must proceed via a review of “relevant” regulations and administrative records, must seek the views of “key” (seems to mean a handful of) informants, etc. These over-detailed prescriptions can turn even the most vivid and arresting conclusions to pabulum.

By 2009 the evaluation function at the department was well staffed and reported through the chief audit and evaluation executive to the deputy minister. Since April of that year, Treasury Board policy has required that all programs be evaluated on a five-year cycle, as well as in advance of periodic re-funding decisions, and that these evaluations be published on departmental web sites. Each department is given a public MAF score on its conduct of the

evaluation function. Each department includes in its departmental Reports on Plans and Priorities its planned internal audits and evaluations for the next three fiscal years, which can be seen on the Treasury Board web site.⁷ DIAND lists 34 planned evaluations ranging from very small programs such as the Inuit Art Foundation to much larger programs such as Child and Family Services. The plan for the evaluation of these programs is coordinated with the plan for the internal audit of the same programs so as to maximize synergies and minimize duplication. The branch budget for 2010-11 was \$6 million devoted roughly evenly to salaries and for contracts with external consultants. The system is adequately resourced and fully bureaucratized.

VALUE FOR MONEY IN FEDERAL OVERSIGHT AND ASSESSMENT

In 2009-10, across the federal government, \$78 million and 515 person-years were devoted to the evaluation function, an increase of more than 100 percent since 2004-05 (Treasury Board Secretariat 2011). The companion program, internal audit, is more backward-looking, but both are seen as essential components of the “assurance” function. The budget for DIAND’s Internal Audit function was comparable at \$6 million, not including the budget of \$4 million for the Assessment and Investigation Branch which pursues allegations of malfeasance. If the balance of resources between evaluation and internal audit is approximately the same across the government, this implies an annual expenditure approaching \$200 million on evaluation and internal audit. To this could be added another \$100 million to operate the Management Accountability Framework (MAF) exercise and the two parliamentary reporting vehicles, the Report on Plans and Priorities (RPPs) and the Departmental Performance Reports (DPRs). We think it would be safe to almost double these numbers to include the indirect costs of other public servants who take time from their regular work to provide information for these oversight

mechanisms. We estimate that the federal government spends at least half a billion dollars a year⁸ on its audit, evaluation and Parliamentary reporting functions.

AN ALTERNATIVE PERSPECTIVE

Before leaving Ottawa, Dobell had thought through and spoken about most of the conundrums and crucial distinctions needed to perform program evaluation: the purpose of the evaluation (forward-looking for purposes of improving designs and allocating resources or backward-looking for purposes of reinforcing accountability relationships); the audience for the evaluation (principal or agent⁹); and the ethical dilemmas associated with risk and decision-making by public managers.¹⁰

The matters of purpose and audience are basic. Hartle frequently observed that asking the public service manager to subject his operations to recurrent comprehensive evaluation is like asking a dog to carry the stick with which she or he is to be beaten (Dobell and Zussman 1981), and Dobell and Zussman note that “the process of policy analysis (including policy and program appraisal, or evaluation) is subject to both procedural impediments, arising out of the fact that the work takes place in an organizational and political context, and to analytical limits arising out of the lack of analytical criteria or relevant information to guide the key choices to be faced” (404) and that:

[There is an] extensive literature on the importance of bureaucratic games, formal and informal pay-off rules or incentive systems, procedural constraints leading to distortions in collective decision processes, and so on. The point is simply that evaluation takes place within a political and organizational context which drives analysis and analysts to an essentially

adversarial role...Within such a framework of advocacy, the bureaucratic incentives do not press in the direction of continuing searching evaluation. (Dobell and Zussman 1981, 413)

The essential distinctions between various purposes and audiences for evaluation are simply not acknowledged in the current federal evaluation and performance measurement policies, which seem to assume that the same set of measures and techniques can serve the needs of program management, resource allocation and accountability. Indeed, the April 1, 2009 *Policy on Evaluation* holds that one flavour of evaluation is to serve all three purposes and audiences:

3.1 In the Government of Canada, evaluation is the systematic collection and analysis of evidence on the outcomes of programs to make judgments about their relevance, performance and alternative ways to deliver them or to achieve the same results.

3.2 Evaluation provides Canadians, Parliamentarians, Ministers, central agencies and deputy heads an evidence-based, neutral assessment of the value for money, i.e. relevance and performance, of federal government programs. Evaluation:

- a. supports accountability to Parliament and Canadians by helping the government to credibly report on the results achieved with resources invested in programs;
- b. informs government decisions on resource allocation and reallocation by:
 - i. supporting strategic reviews of existing program spending, to help Ministers understand the ongoing relevance and performance of existing programs;

- ii. providing objective information to help Ministers understand how new spending proposals fit with existing programs, identify synergies and avoid wasteful duplication;
- c. supports deputy heads in managing for results by informing them about whether their programs are producing the outcomes that they were designed to produce, at an affordable cost; and,
- d. supports policy and program improvements by helping to identify lessons learned and best practices (Treasury Board Secretariat 2009).

This across-the-board philosophy has been questioned by seasoned practitioners. Writing in the *Canadian Journal of Program Evaluation*, Greg Mason worries that:

The federal model, as typified by what appears to be the rote methodology (consisting of document/file review, literature review, interviews, focus groups and case studies), masquerades as truth-seeking under the banner of multiple lines of evidence. It is also apparent that evaluation has drifted close to becoming a form of audit, and the standards of validity reflect a dangerous naivety. (Gauthier et al 2009, 10)

We have noted elsewhere (Clark and Swain 2005) that the propensity of federal central agencies (a) to ignore basic human realities in government and (b) to resort to across-the-board, template-driven policies, frequently leads to demands that appear surreal to departmental managers. As a current example, here is the admonition in the current *Performance Reporting Guidelines* encouraging departments to include honest self-evaluations in their annual Departmental Performance Reports to Parliament:

Report positive and negative aspects of performance

Balanced reporting enhances the credibility of reporting. Moreover, reporting on both the positive and negative aspects of performance meets the intention of public performance reporting—to provide the necessary information for scrutiny and decision making. Telling only half of the story is not useful.

Performance information is not fairly presented when the information is limited to successes and minimizes, or even avoids, discussion of matters that did not unfold as planned. In order for reporting to be fair, key information must not be omitted. When discussing results achieved, departments should discuss what went according to plan as well as cases where things did not go according to plan but risks were mitigated. Departments should also discuss results that were not achieved, noting how and why plans were not implemented as intended. It is also important to include explanations of how the department uses both positive and negative results to make adjustments and improvements toward achieving its strategic outcomes. Findings that emerge from evaluations are important sources of information in framing these discussions (Treasury Board Secretariat 2010, 30).

Dobell had painstakingly analyzed and then explained to his central agency colleagues forty years ago why the personal and organizational dynamics in government make balanced self-evaluation impossible. Recognizing well-researched human and bureaucratic realities, Dobell and Zussman (1981) asserted that:

Decision support systems for good management of an agency are different from mechanisms for political accountability to Parliament and the public.

Evaluation procedures modelled on integrated financial management systems and based upon efficiency in resource use are essential for the first, and (almost) useless for the second. (Dobell and Zussman 1981, 418)

Because even a very good evaluation system cannot discharge the government's responsibility for accountability to Parliament and the public,

...our answer to the question 'Who should do these evaluations?' is that in the case of summative evaluation (designed to answer the question 'How well are we responding to the problem?') we should all do it – academics, journalists, parliamentary staff, the community at large – much more than we do. We should do it case by-case, randomly, as events arise, but on the basis of full and open access to the relevant information. (Dobell and Zussman 1981, 422)

HOW EVALUATION MIGHT HAVE EVOLVED

Now for our thought experiment. If Rod Dobell had held prominent positions in the Treasury Board and Indian Affairs during the last three decades would this half billion dollars a year be spent differently? Would the Treasury Board Secretariat, perhaps in conjunction with other central agencies, be taking a stronger role in forward-looking policy evaluations on crucial national issues such as those pertaining to the health and education of Canada's aboriginal population? Would the department's program evaluations look as they do today? Would a different approach to evaluation have caused policies and programs for aboriginal peoples to evolve in more positive ways?

We believe that if Dobell had remained in charge of the Government of Canada's evaluation policy the resulting studies would be more useful and more realistic than they are today. We would have seen more of the government's evaluation resources directed to places where they could make the most difference, with some of the bigger studies godfathered by the centre, particularly where resource allocation and instrument redesign were major concerns. And we think that we would have seen more effort to engage leading scholars from universities, and to make more program information public to facilitate "evaluation by all of us."

Dobell kept thinking about these issues through the 1980s and 1990s as political scientists elaborated their post-modernist theories and the analytical premises of the Planning Branch era came under intense critical scrutiny. He provided his most recent synthesis in a lengthy research paper for Ontario entitled "The Role of Government and the Government's Role in Evaluating Government: Insider Information and Outsider Beliefs." He writes:

There are distinct approaches to project evaluation, program evaluation, performance evaluation, and policy evaluation, running from the most 'factual' and amenable to uncontested characterization to least 'factual' and most subject to conflicting and contested perspectives...For the former, where facts, figures of arithmetic, and technical or engineering considerations dominate, accounting and audit can support informative reporting. For the latter, one is left with rhetoric and figures of speech, with only a process of narrative, dialogue, and deliberation to rely upon to achieve convergence within what one hopes will be accepted as a legitimate communicative forum. (Dobell 2003, 49-51)

And he closes his research report with:

The pursuit of equal opportunity for all, for the flourishing of human potential, can only be pursued, it may be argued, through the self-determination made possible within the institutions and norms of a knowledge-based democracy and an experimenting (learning) society. It is this vision, not a model of formal accountability, that should underlie the evolution of evaluation activity in Ontario over the coming decades.

(Dobell 2003, 53)

All this may be conceding a trifle too much to the contingent and the rhetorical, but it certainly is inconsistent with the formulaic, rules-based evaluations currently in vogue in Ottawa.

Now let us turn to how Dobell's presence might have influenced the conduct of evaluations in at DIAND.

In the first place, we would not expect to see a rigid adherence to a TBS-enforced table of contents, nor the use of TBS-mandated jargon. Dobell has long argued that a narrative specific to the program and the audience lends nuance and readability. As Secretary he would not have countenanced the more mindless parts of proposed central agency guidance; had he received it, as a line deputy minister, he would in the most graceful, witty but unanswerable way ignored it. Second, in either role he would likely have insisted on a broad and forward-looking component to every evaluation. Questions of relevance, instruments, alternatives, and unspoken as well as official objectives would have always been part of the remit; narrow studies such as those reported in INAC (2008) would have been fewer and less formulaic. There would have been more weight, and more resources, devoted to a smaller number of evaluations of really important

programs, rather than a lock-step, cookie-cutter approach to trivial matters just because TBS, having sunsetted most programming, requires even the smallest, most immaterial activity to be formally evaluated before new funding is granted. The crude central agency conception of “risk,” as a basis for all planning and reporting, would not long survive in the domain of the author of “Social risk, political rationality, and official responsibility: risk management in context” (Dobell 2002). A much more nuanced view, not so overwhelmingly centred on financial or reputational risk to the federal government, could have been expected.

Finally, we would have expected, for those major studies that evaluated really important programs, to have seen much wider participation. Aboriginal stakeholders and critics, engaged academics, even the press and the general public, would have been urged to take part from the beginning – from the day when the scope of the study and the objectives of the program were first examined. In this respect Justice Dennis O’Connor’s organization of the Walkerton Inquiry, to which Dobell was an advisor, is instructive. Part 1 had a backward-looking focus on how that particular tragedy happened. The proceedings were court-like, with evidence widely sought and subject to cross-examination by the numerous parties with standing, the legal version of an audit. Part 2 used that evidence, and much more, to make recommendations for the future. Its proceedings were entirely different except in the Commissioner’s insistence that all the work be done in public in both parts. Stakeholders of all sorts were invited to present and comment on all important topics: the only rule was they had to do so in public, in front of parties who might have had quite different views.¹¹ At DIAND, Dobell would have encountered the unwillingness of many aboriginal stakeholders to take part in official evaluations, but to a much lesser degree than afflicts present practice as there would have been radically fewer calls for participation, on vastly more important topics.

WHAT A DOBELL-STYLE FORWARD-LOOKING EVALUATION MIGHT LOOK LIKE

We have been impressed with how forward-looking the Dobell conception of evaluation studies was. Even in the 1970s, Cabinet evaluation studies were not primarily intended to verify whether government programs and their managers had lived up to previous commitments; they were searches for better solutions to the big policy problems at hand. This was “evaluation for improvement.” It did not pretend to be “evaluation for accountability.” We think that, if Dobell had had his way a substantial portion of the treasure now being spent on evaluation would be devoted to such studies. What might such evaluation studies look like in aboriginal affairs?

Below we provide a speculative example based on recent experience with the work of the Expert Panel on Safe Drinking Water for First Nations (INAC 2006).

Let us imagine an evaluation study on aboriginal housing. This study would involve substantial consultation with aboriginal communities and would include First Nations experts from outside government on the study team. The outline of the final report might look like:

Outline for a Dobell-style evaluation study entitled Policy Options for Improving the Quality and Cost-effectiveness of Support for Aboriginal Housing

- *The case for reform.* This section would document the sad and overcrowded state of on-reserve housing despite the considerable, often ineffective, federal investment, noting the connections to related domains such as health, education and incomes.
- *What is good housing and how is it achieved.* This section would examine good models from Canada and abroad, surveying a large literature not just on housing standards but

on successful ways of financing and maintaining housing stock once constructed, taking into account the special hurdles faced by isolated small communities.

- *Challenges and complexities.* This section would examine real housing on the ground in real Canadian reserve communities. Problems and their causes, ranging from unavailability of capital to poor maintenance practices and incentives would be analyzed. The circular causes of poverty, poor housing, poor health and poor education would not escape notice. Constraints on achieving good housing, ranging from the Indian Act to capital availability, would receive particular attention.
- *What we heard.* The 2006 drinking water study included public hearings in or near reserve communities in ten regions across the country as its principal means of involving the consumers and providers of drinking water on reserves. A study of housing might well adopt this technique as a means of escaping, if briefly, the clutches of distant experts and those with axes to grind. The study team, which for the reasons cited above would be composed not of public servants but supported by them, would report on insights gained in the field, reconciling them where necessary with official statistics and views.
- *What would be supported and how.* Drawing on the evidence, factors leading to better housing outcomes in small places would be isolated and commented on, with distinctions made between what governments can do and what they cannot. The section would conclude with a short list of critical elements, together with their relation to broader community factors, on which government action might have some lasting effect.

- *Program design options.* This section would package critical elements and the relaxation of constraints in several different ways. With each would be associated a probability, and the costs (not just the financial ones), of achieving a particular level of the objective of good housing for all.

This hypothetical example shows that dramatically different uses of evaluation resources are possible. This kind of evaluation would be controversial and require much greater engagement from senior officials and ministers. The effort would likely need to be led by a person of force, sophistication and wit, someone not afraid to raise his head above the parapet of civil service anonymity. No one could be better than the writer of the 1989 masterpiece, who concluded that

Even in a world of profound uncertainty and missing data, explicit and formal analysis offers the essential decision tool for informing and guiding the public manager in the often agonizingly difficult decisions that must be taken on matters of public risk. Within a framework of principle, analysis does help. (Dobell 1989, 10)

CONCLUSIONS AND SUGGESTIONS

Is it possible to draw conclusions from this review of evaluation in aboriginal affairs and our imagination of a Dobell-driven approach to policy analysis? Are there better ways for the Canadian government to spend the \$78 million per year currently devoted to evaluation-related activity? We believe so.

It is not realistic to suggest that the evaluation function be managed by people as highly trained, analytically rigorous and intellectually curious as Rod Dobell. The world has few such people. But here are four suggestions that we believe ordinary mortals could apply.

1. The government should make explicit the distinction between “evaluation for improvement” and “evaluation for accountability” and devote the vast majority of evaluation resources to the former.
2. Senior officials should apply more judgement in selecting topics for evaluation. Rather than applying across-the-board rules such as “every program, every five years” or “before seeking renewal of any program of specific duration,” the topics should be selected on the basis of *the extent to which evaluation might lead the government to make material changes in program design or funding*.
3. The central agencies that advise ministers on policy priorities and funding should become more involved in the selection and design of evaluation projects.
4. There should be fewer, but more thorough, evaluations and they should draw on the techniques of analysis employed in academic social sciences research and the techniques for engaging interested parties employed by successful public enquiries.

We think these modest suggestions are fully in the spirit of Rod Dobell.

¹ Since 2011, the government refers to the department as Aboriginal Affairs and Northern Development (AANDC). To avoid confusion, we use the name used throughout the period under review, which remains its statutory name today.

² In our look at the underlying theory we can think of *evaluation* and *policy analysis* almost interchangeably since the fundamental skills and assumptions of high-end evaluation are so similar to those for general policy analysis.

³ An exceptional number of these recruits went on to the most senior posts in the public service in the decades to come, including Mel Cappe, Barry Carin, Ian Clark, Frank Claydon, Mark Daniels, Rod Dobell, John Edwards, Len Good, François Lacasse, Maurice Laffontaine, Paul-Henri Lapointe, Bruce Montador, Russ Robinson, Michael Wolfson, and David Zussman.

⁴ One of them, on industrial programs and subtitled “Giving with Both Hands,” set an important but unfollowed precedent by looking at tax expenditures with the same optic as direct expenditures. To this day TBS does not require the Department of Finance to evaluate the effectiveness, costs and benefits, or economy of tax expenditures.

⁵ A 1991 central agency edict to cap expenditures on post-secondary education, one of the Department’s most successful programs, was met by nationwide protests and the shuffle of the line minister who tried but failed to defend an unthinking instruction from the center: a good example of a decision taken in the teeth of evaluation evidence.

⁶ Including—full disclosure—the present authors.

⁷ At <http://www.tbs-sct.gc.ca/rpp/2011-2012/info/ia-vi-eng.asp#ian> (accessed 16 August 2011).

⁸ Official numbers always underestimate the hidden costs associated with these oversight functions. On the cost of the MAF, Lindquist (2009) writes: “One way to focus our attention on the implications of MAF is to consider its annualized costs. As a starter, let’s consider a very conservative and “back-of-the-envelope” estimate: 10 FTEs in the MAF directorate and its parent unit; 60 FTEs associated with the program sectors and policy centres of TBS and related

central agencies (20% of time of around 300 full-time staff); 350 FTEs in departments and agencies (say 6 FTEs on average for the 55-60 departments and agencies covered by MAF each year, recognizing the larger departments devote more resources than smaller agencies). This leads to a conservative estimate of 420 FTEs and an annual cost of \$42 million (priced at \$100,000 per FTE) to run MAF. There are many officials who would argue that this grossly underestimates the cost of staff time in central agencies, departments, and agencies. While this estimate does not include the information technology and management requirements, the costs may be lower than it might seem since MAF relies on data and information already in the system, provided as part of a much larger system of accountability and reporting.” We would suggest that the cost of running the RPP and DPR processes are similar.

⁹ Dobell articulated this in terms of the chain of delegation between the party who has delegated, and another who has accepted, certain responsibilities (see Dobell 1975, 5-6). Today we would likely use the language of principal-agent theory.

¹⁰ Dobell later brought together these latter ideas with the then fashionable idea that public servants needed to be more entrepreneurial in a brilliant article in the *American Review of Public Administration* entitled “The Public Administrator: God? Or Entrepreneur? Or are they the Same in the Public Service.” (Dobell 1989). The article opens with: “We all know that to innovate is great. But we also know that to err is human, and to forgive may be divine – but it is not customary. A tradition of entrepreneurship can only flourish in a culture in which the costs of error are not so great as to demand divine dispensation on a continuing basis.” These views were further elaborated in Dobell (2002).

¹¹ The result was interesting: a pleasing degree of civility, an unwillingness to use intemperate language, and the discovery that areas of agreement far outnumbered the others.