

# Linking Evaluation to Expenditure Reviews: Neither Realistic nor a Good Idea

---

John Mayne  
Ottawa, Ontario

**Abstract:** *It is frequently assumed and not contested that evaluation should play a significant role in budgeting and, more specifically, in expenditure reviews. This article argues otherwise: that evaluation is neither fit nor designed to play such a role. Rather, if there is a desire by budget officials for credible evidence on the performance of interventions, then they need to invest in a different form of evaluation, namely, expenditure evaluations, separate and distinct from ministry-based evaluation.*

**Keywords:** *evaluation challenges, expenditure evaluation, expenditure reviews*

**Résumé :** *On tient souvent pour acquis que l'évaluation devrait jouer un rôle important dans les décisions et les exercices de révision budgétaire. Ce rôle est rarement contesté et est souvent perçu comme l'un des objectifs de l'évaluation. Dans cet article, nous faisons l'argument inverse : l'évaluation n'est ni adaptée ni conçue pour jouer un tel rôle. S'il y a un besoin d'intégrer aux décisions budgétaires de l'information crédible sur la performance des interventions, alors nous devons investir dans une autre forme d'évaluation, soit l'évaluation des dépenses, une fonction qui devra être séparée et distincte de l'évaluation institutionnalisée dans les Ministères.*

**Mots clés :** *défis de l'évaluation, évaluation des dépenses, revue des dépenses*

This special issue is addressing the possibilities and challenges of evaluation playing a significant role in expenditure reviews undertaken by central units such as ministries of finance or budget offices, focused on reducing or reallocating government expenditures. By *expenditure reviews* I mean to include both ongoing program review as well as one-shot reviews of expenditures, both of which occur in the Canadian federal government (Dumaine, 2012) and elsewhere (Robinson, 2013).

Robinson (2014) argues that for expenditure reviews to use evidence from evaluation effectively, both such reviews and evaluation need restructuring. I agree. The focus here is on the evaluation side of the equation rather than on expenditure reviews per se, which are discussed elsewhere in this special issue. And, further, my focus will be mostly on evaluation generally, rather than on a program evaluation function in a particular government. Robinson (2014) discusses the

**Corresponding author:** John Mayne, 1607, 3580 Rivergate Way, Ottawa, ON, K1V 1V5; [john.mayne@rogers.com](mailto:john.mayne@rogers.com)

shortcomings of government-wide evaluation systems in expenditure reviews. [Shepherd \(2018\)](#) discusses the specific case of Canada.

The issue then is whether and, if so, how evaluation could support expenditure reviews. And the evidence from the past 30–40 years is not encouraging. History suggests evaluation has not played a significant role in expenditure reviews, either in Canada or elsewhere ([Dumaine, 2012](#); [Robinson, 2014](#)). This history is often conveniently forgotten; the belief still seems to be that we just need better evaluations, better forms of expenditure review, and better processes for linking evaluation into expenditure management. I think that belief is mistaken; there are more fundamental problems.

First I discuss the promise of evaluation playing a significant role in expenditure reviews, why it seems reasonable to expect evaluation to be able to contribute to expenditure reviews. This is followed by a discussion of the reality, outlining why such a role is not only not practical but seriously flawed conceptually. The article concludes with an exploration of the implications and with suggestions for linking a revised form of evaluation with expenditure reviews.

## THE EXPECTATION

The idea to link evaluation and expenditure reviews, and resource allocation more generally, is seemingly simple and one that has been around as long as has evaluation. Evaluation aims to assess the performance and worth of programs—that is, whether programs are working or not—and hence should be of significant use in determining expenditure allocations. Programs that are shown to be not working, that are ineffective, should be terminated or cut back.

In 1981, the government of Canada made it clear that

[the purpose of program evaluation] should be to periodically produce credible, timely, useful and objective findings on programs appropriate for resource allocation, program improvement and accountability. ([Treasury Board Secretariat, 1981](#), p. 3)

The simple and apparently convincing logic behind this belief continues till today; many governments continue to expect evaluation to be able to play such multi-purpose roles. [Robinson \(2014\)](#) in his review of connecting evaluation to budgeting, while arguing the need for more relevant evaluation, notes that “it is unambiguously clear at the theoretical level that evaluation is essential to performance budgeting” (p. 23). In my view, it is not unambiguously clear at all.

I had a role in articulating the above 1981 stated purposes for evaluation in the government of Canada. At the time it seemed reasonable, logical, and quite doable. However, this belief was not based on any concrete evidence of evaluation actually playing such a role in expenditure management. It just seemed obvious and not contested in the literature. In hindsight, we were naive, promising much too much for evaluation to deliver. We were not basing decisions on what evaluation aims to deliver: evidence.

## THE REALITY

As noted, there has not been a lot of evidence of evaluation playing much of a role in expenditure reviews—quite the contrary. If that is the experienced evidence, what is the explanation? Robinson has discussed some of the explanations, which I summarize:

- Evaluations produced by evaluation units within ministries tend to be process evaluations and focus on operational issues, not producing the kind of information useful for budgeting. (2014, p. 23)
- Even when managed by central government units, such as in Chile (2014, p. 25), the evaluations tend not to focus on issues pertinent to expenditure reviews.
- Timing is usually a problem. Evaluation assessing the impact of interventions take time and often experience delays, making it difficult for findings to be available when needed for expenditure reviews (2013, p. 27). And when they are available, of necessity they may be several years old.
- The quality of evaluations is often suspect by budget officials (2014, p. 23).

And the solutions suggested focus on making the evaluations more credible and relevant for expenditure reviews, and on making expenditure reviews more focused on performance information. In essence, the suggestions are to improve demand and enhance the supply to meet that demand.

The points made in these assessments of evaluation and expenditure reviews are largely valid. However, in my view, they do not go far enough. The problem is more fundamental; the idea of linking evaluation to expenditure reviews is conceptually flawed.

What does not seem to have been seriously considered is

- that evaluations often try to serve several masters,
- what the “performance” of a program typically entails,
- the type of information evaluation produces, and
- the type of performance information budget officials appear to want.

### *The Clients for Evaluation*

The government of Canada’s 1981 purpose statement quoted above suggests three purposes to be served: resource allocation, program improvement, and accountability. Evaluations done solely for a program or within a ministry might try to serve these various ends, albeit with difficulty. But if those three purposes are associated with three different parties—namely, budget officials, program managers, and Parliament respectively—then a whole different issue presents itself. Indeed, a single evaluation or even an evaluation function cannot serve well these three quite different clients. If they try to do so, the likely result is that all three clients will be unhappy with what is delivered (Mayne, Divorski, & Lemaire, 1999; Robinson, 2018).

In the Canadian federal system, the clients for the evaluations are the deputy heads of the ministries. The reality is that it is they for whom the evaluators work, and it is they who, with the senior management team, decide the resources, scope, focus, and timing of evaluations. To imagine that these evaluations can also serve budget officials and perhaps Parliament is quite unrealistic. Evaluations can realistically only well serve in a credible manner one client at a time.

### ***What Program Performance Entails***

Public sector programs are interventions in the social and economic lives of people and businesses, attempting to change behaviours. As such, there is often considerable uncertainty involved in exactly how, or indeed if, they will work. They are often dealing with complex phenomena, with many interrelated components, involving a variety of private and public actors. Some aspects of a program may be working fine, others not so, and the various components are connected. This makes thinking about and, of course, evaluating performance a challenge.

Performance in the form of effectiveness is usually taken to mean the program is meeting its objectives. But, as I argue elsewhere (Mayne, *in press*), just what “meeting objectives” means is not at all clear. There are often several factors contributing to bring about objectives (results) in addition to the program itself. So what does effectiveness really mean in such cases? In practice, the program may be contributing to the objectives along with other causal factors. And in so doing it may be playing a significant role or a more modest role. But, on its own, it is unlikely to be causing the observed results. So, what does “meeting its objectives” mean in practice?

All this is to say that even ignoring measurement and attribution issues—which can be very significant—the concept of performance of a program is often not straightforward, and findings on performance need to be thoughtfully interpreted. It is also the case, I would argue, that most public sector programs are a mixed bag in terms of performance, rarely completely effective or completely ineffective. As a result, the performance of programs is most often much more nuanced than a simple yes–no rating.

### ***What Evaluation Delivers***

Evaluation assesses the performance of a program. But given the nature of programs and their context, and the complexity of what performance entails, the assessments are rarely simple findings that the program is or is not performing. Good evaluation tells a performance story, about what is working, for whom, and why, and what is not working and why. And the findings are usually caveated, not providing 100% proof on many issues, and hence may be contested.

A key reason for this is the challenge of assessing causality, of demonstrating a link between what the program is doing and any observed results. This assessment is essential, since observed effects may be the result of other non-program actions, such things as other programs, markets, or trends in social-economic factors. If such is the case, then the program is likely not performing very well, even if the expected results from the program are being realized. Evaluation tries to sort out these

uncertainties and complexities, and to provide useful insights on how and why the program is performing, but often it often leads to nuanced and caveated findings.

All this complexity means that evaluations take expertise, time, and resources. They can credibly deliver an assessment of performance, but not simple, black or white findings.

### ***The Purpose of Evaluations***

Evaluations are undertaken for different purposes. The 1981 quote above from the Canadian government noted three typical purposes:

- for resource allocation, such as expenditure reviews,
- for program improvement and learning, and
- for accountability.

There is debate in the evaluation literature on the extent to which a single evaluation or evaluation function can indeed meet these different purposes ([Hutchings, 2014](#); [Regeer, de Wildt-Liesveld, van Mierlo, & Bunders, 2016](#)). I have argued that meeting accountability and learning aims in a single evaluation is not practical ([Mayne, 2007, 2015](#)). While there are calls for evaluations to address both accountability and learning ([Aucoin, 2000](#); [Guijt, 2010](#); [Hutchings, 2014](#)), in practice an accountability focus tends to push out learning ([Guijt, 2010](#); [Perrin, 2015](#); [Regeer et al., 2016](#)). Hence the call by some in Canada for an “evaluator general” to complement the auditor general and deliver accountability reports to Parliament ([D’Aloisio et al., 2007](#)). Similarly, I would argue, and as noted earlier, that seeing a single evaluation meeting all three purposes is unrealistic.

### ***What Budget Officials Want***

[Robinson \(2014\)](#) suggests questions that budget officials want answers to. In particular,

Which existing programs (or components of programs) should be terminated because they are ineffective or not sufficiently effective to justify their cost? (p. 11)

Note that this is not simply a question of identifying ineffective programs, but also of establishing whether any such programs can realistically be fixed. It makes no sense to cut the funding of an ineffective program or service if it could be made effective at reasonable cost by modifying policy design or implementation strategy.

But, in many cases, I would argue that this question is not that meaningful. As noted, few programs will be shown to be ineffective in the yes–no sense the question is looking for ([Robinson, 2018](#)). And justifying performance in relation to the costs of the program (even using simple expenditure as “cost”) is again a quite complicated question with significant subjective values being used. Indeed, as [Robinson \(2018\)](#) argues, usually such cost-effectiveness assessments are not

done by evaluators (I do not see evaluators having any special skills in costing) and in practice are usually made informally.

This is all to say that while budget officials may indeed want to have nice and neat evidence-based answers about the worth of a program,<sup>1</sup> in practice such questions are not that meaningful or forthcoming. There often is not a bottom line rate of return. As we saw above, the performance of many programs is a much more complex and nuanced concept. Budget officials need to accept this and learn to thoughtfully interpret the kind of information they do get from evaluation activity.

## SO WHERE ARE WE AND WHAT CAN BE DONE?

The fact, then, is that evaluation operates in the real world of uncertainty and complexities, leading to findings on the performance of programs that inherently must be nuanced and caveated to accurately tell the performance story. This kind of information can still *inform* resourcing decisions but doesn't lead to mechanistic implications about funding. *Evidence-based* decision-making is unrealistic; aiming for *evidence-informed* decision-making is the most one can hope for. A program may be weakly performing because it was inadequately resourced. A well performing program may be so because the problem it aimed to address no longer is a problem.

On a more positive note, it seems to me that there are ways to better link “evaluation” with expenditure reviews, which I discuss below. These build on suggestions by [Robinson \(2014\)](#) and others.

### *Using the Decentralized Evaluation Function*

Much evaluation in governments is decentralized to spending ministries where it plays a very useful role serving senior management in adjusting programs to make them most effective and more cost-effective. It can also provide input into resource reallocation at the ministry level and help ministries account for their use of public funds. These are valuable and traditional roles for evaluation. When a program is found to be not performing very well, this usually triggers changes to fix the program. What an internal evaluation function cannot do very well is seriously question the existence of its own organization's programs. Given where the evaluation is located, it is unrealistic to expect otherwise. As [Robinson \(2014\)](#) notes, “the notion that a single government-wide evaluation system can serve both of these purposes [budgeting vs. management and policy] effectively should be reconsidered” (p. 6). This is indeed what I and my co-authors argued 18 years ago (Mayne et al., 1999). Different evaluation purposes require different types of evaluations positioned in different locations within a jurisdiction.

From all the evidence, experience over the last 30–40 years, and logic, I would argue that government evaluation systems decentralized to ministries—such as in the Canadian federal government—should not be expected to significantly contribute to central expenditure reviews. Such an expectation is not only unrealistic

and wishful thinking, but trying to push decentralized evaluation in that direction undermines what it does well: credibly serve ministries in regularly assessing its programs so they can be adjusted as needed—including sometimes their resources being used elsewhere in the ministry.

I am not suggesting that such evaluation should be off-limits to expenditure reviews. They should, after all, be publically available. Rather, there is a need to be modest as to what useful information for budget decisions is likely to be found in such evaluations, recognizing the timing problems, the complexity of program performance, and the scope of ministry evaluations. Some evaluations may indeed raise red flags that the expenditure review needs to pursue, or raise obvious questions about what the ministry has done about observed weak performance. And, of course, they might be used to help justify expenditure review decisions already made—the symbolic use of evaluation (Patton, 2008).

### ***Build a Central Evaluation Capability***

One obvious logical suggestion is to build a central agency evaluation capacity, focused on expenditure management. Robinson (2014) argues this case:

To ensure that evaluation meets the needs of budgeting, it is therefore necessary that the Ministry of Finance (possibly in association with other relevant central agencies and with relevant input from the political leadership) control and manage a targeted schedule of evaluations of baseline expenditure that is specifically designed to support budget preparation. . . .

The timing of the budgetary evaluation cycle must also be synchronized with the budget preparation process. This means that:

- o Evaluation reports must be supplied to the ministry of finance at right time of the year (early in the budget preparation process) to be fully taken into account in the budget.
- o Evaluations must be carried out quickly, so they do not lose pertinence. . . . Different time limits should be set for different types of evaluation. For example, a rapid evaluation (of which the core element is a program logic analysis) can be carried out within, say, 4 months. A longer time limit—perhaps 18 months—needs to be set for outcome evaluations. The timeframe for process evaluations comes between these.
- o The budgetary evaluation cycle must reflect the timing of the budget preparation process—annual or multi-annual. In most countries this means that evaluation results can be fed into the budget preparation process every year. In countries where ministry budgets are set for several years (for example, the United Kingdom), evaluation results will serve as inputs to budgeting at the relevant time interval (e.g., every three years). (p. 35)

A central evaluation unit undertaking selective and strategic evaluations on an ongoing basis seems an obvious and necessary step toward making better use of evaluation thinking and approaches. One obvious advantage is that such central

evaluations could focus on horizontal programming themes and government policies across ministries.

### ***Develop a New Form of Evaluation***

The idea of using what Robinson calls program logic analysis in the form of rapid evaluation is promising. This is essentially a form of *theory of change analysis* (Mayne, 2017) where considerable insight on a program can often be gained assessing the viability of the theory of change of a program and using existing data.

But I would go further and argue that such centrally managed *expenditure evaluations* be seen as a different form of evaluation, one that has not received much attention in the literature. They could involve a number of elements:

- In many cases, a short time frame, focused exclusively on expenditure-related issues.
- The client would be the budget officials and the evaluations would not be aimed at trying to improve programs, although that could indeed be an outcome through subsequent actions taken by a ministry.
- Undertaking rationale analysis to confirm alignment with current government priorities.
- Making use of prior evaluations, such as from ministries, and of related studies to the extent feasible, looking, for example, for situations where previously identified program shortcomings have not been addressed.
- Carrying out theory of change analysis, akin to program logic analysis. In particular, given the time frames, it would often assess the likelihood that a program is contributing in a meaningful manner to desired outcomes, rather than trying to confirm the causal links.
- Using theory-based approaches (Treasury Board Secretariat, 2012) to conclude on causality, rather than using more traditional experimental approaches.
- Where expenditure reviews can be planned to inform future budget decisions, such as in the UK, the longer time frame could be used to undertake more in-depth evaluations of government programming areas and issues.
- Producing relevant information to inform upcoming budget decisions, and hence require budget officials knowledgeable and competent in using the kinds of caveated findings the expenditure evaluations would deliver.

These expenditure evaluations could be complemented with efficiency studies, done as a separate exercise.

Expenditure evaluations would, of course, require resources, resources over and above those spent on regular evaluation in ministries. If budget officials and ultimately politicians want to have better information and evidence on how well their programs are doing, they need to be willing to invest resources to get it. One

would expect expenditure evaluations to show their worth. One would also hope that such resources would not be taken from existing evaluation efforts.

Some experimenting with this new form of evaluation would be necessary, to see what can be produced and used by budget officials, and what kind of principles would support such evaluations. They would, for instance, unlikely be participatory in nature or gather extensive amounts of new data. Hopefully, if this approach were adopted, a literature on practice would slowly grow.

Undertaking these evaluations could be challenging. As Robinson suggests, they would probably be carried out by external consultants managed by the budget officials. But they would need access to data and information in the ministries, and have some level of co-operation with officials and evaluators in the spending ministries. While ministries might be apprehensive and try to play games on data and access, they might, on the other hand, see advantages to be reasonably co-operative, so as to be able to influence, at least to some extent, the expenditure evaluations. The central evaluation unit would need access to data and documentation such as is available to state audit offices.

## CONCLUDING REMARKS

The experience of trying to link regular evaluation with expenditure management is not great. The explanations for why this is so are many and well documented. I have added to the explanation, suggesting regular evaluation for the most part doesn't produce the kind of information that helps much in budget decisions. We should draw the obvious conclusion: it can't be done. Furthermore, it shouldn't be done—one tries to do so at the expense of what regular evaluation does well, namely, serve ministries.

This article has discussed the reasons for this conclusion:

- Evaluation cannot serve multiple masters. Evaluations in ministries serves ministries well and needs to continue to do so. Confounding those evaluations with the interests of budget officials weakens them and will not serve budget officials well.
- Evaluation does not produce neat simple answers about the performance of most programs; program performance is a much more complex concept. Budget officials looking for findings from evaluation that show a program is or is not effective are being unrealistic and perhaps disingenuous.
- Evaluations require time and resources to deliver credible findings on the performance of programs, time often not available in expenditure reviews.

To have evaluation serve expenditure reviews well, I argue the need for a new form of evaluation: expenditure evaluation. Expenditure evaluation would be aimed solely at providing useful information for budgeting decisions and be managed by budget officials.

## NOTE

- 1 Some budget officials may, of course, just want evidence that justifies their budgeting decisions, with evaluation just playing a symbolic role. Robinson's interviews, in 2013, of budget officials in several countries did indicate, however, an interest in solid evidence-based information of programs (Robinson, 2018).

## REFERENCES

- Aucoin, P. (2000). *The public service as a learning organization: Maintaining the momentum in public service reform*. Ottawa, ON: Canadian Centre for Management Development.
- D'Aloisio, G., Laurendeau, M., Neimanis, V., Obrecht, M., Porteous, N., Prieur, P., & Witmer, J. (2007). An evaluator general for Canada—A solution for filling the accountability void? *Canadian Government Executive*, 13(7).
- Dumaine, F. (2012). When one must go: The Canadian experience with Strategic Review and judging program value. *New Directions for Evaluation*, 2012(133), 65–75. <https://doi.org/10.1002/ev.20007>
- Guijt, I. (2010). Accountability and learning: Exploding the myth of incompatibility between accountability and learning. In J. Ubels, N.-A. Acquaye-Baddoo, & A. Fowler (Eds.), *Capacity development in practice* (pp. 277–292). London, UK: Earthscan.
- Hutchings, C. (2014). Balancing accountability and learning: A review of Oxfam GB's global performance framework. *Journal of Development Effectiveness*, 6(4), 425–435. <https://doi.org/10.1080/19439342.2014.971552>
- Mayne, J. (2007). Evaluation for accountability: Reality or myth? In M.-L. Bemelmans-Videc, J. Lonsdale, & B. Perrin (Eds.), *Making accountability work: Dilemmas for evaluation and for audit* (pp. 63–84). New Brunswick, NJ: Transaction.
- Mayne, J. (2015). Structuring evaluation for learning. *Canadian Government Executive*, 21(10), 14–15.
- Mayne, J. (2017). Theory of change analysis: Building robust theories of change. *Canadian Journal of Program Evaluation*, 32(2), 155–173. <https://doi.org/10.3138/cjpe.31122>
- Mayne, J. (in press). Realistic commissioning of impact evaluations: Getting what you ask for? In A. Paulson & M. Palenberg (Eds.), *Evaluation and the pursuit of impact*. Taylor and Francis.
- Mayne, J., Divorski, S., & Lemaire, D. (1999). Locating evaluation: Anchoring evaluation in the executive or the legislature, or both or elsewhere? In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice* (pp. 23–52). New Brunswick, NJ: Transaction.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Perrin, B. (2015). Bringing accountability up to date with the realities of public sector management in the 21st century. *Canadian Public Administration*, 58(1), 183–203. <https://doi.org/10.1111/capa.12107>
- Regeer, B. J., de Wildt-Liesveld, R., van Mierlo, B., & Bunders, J. F. G. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation*, 22(1), 6–28. <https://doi.org/10.1177/1356389015623659>

- Robinson, M. (2013). *Spending Reviews: 34th Annual Meeting of OECD Senior Budget Officials*. Paris, France: Organisation for Economic Co-operation and Development. Retrieved from [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=GOV/PGC/SBO\(2013\)6&doclanguage=en](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=GOV/PGC/SBO(2013)6&doclanguage=en)
- Robinson, M. (2014). *Connecting evaluation and budgeting* Evaluation Capacity Development Working Papers Series Washington, DC: World Bank Independent Evaluation Group.
- Robinson, M. (2018). The role of evaluation in spending review. *Canadian Journal of Program Evaluation*, 32(3), 305–315.
- Shepherd, R. P. (2018). Expenditure reviews and the federal experience: Program evaluation and its contribution to assurance provision. *Canadian Journal of Program Evaluation*, 32(3), 347–370.
- Treasury Board Secretariat. (1981). *Guide on the program evaluation function*. Ottawa, ON: Treasury Board of Canada, Office of the Comptroller General of Canada.
- Treasury Board Secretariat. (2012). *Theory-based approaches to evaluation: Concepts and practices*. Retrieved from <http://www.tbs-sct.gc.ca/cee/tbae-aeat/tbae-aeat-eng.pdf>

## AUTHOR INFORMATION

**John Mayne** is an independent advisor on public sector performance. Over the past 13 years he has focused on international development evaluation and RBM. He has authored numerous articles and reports, including several on contribution analysis, and co-edited eight books on program evaluation. In 1989 and in 1995, he was awarded the Canadian Evaluation Society's Award for Contribution to Evaluation in Canada. In 2006, he was made a Canadian Evaluation Society Fellow. Dr. Mayne's current interests are approaches for strengthening impact evaluation and useful theories of change in complex settings.