

EVALUATION CAPACITY DEVELOPMENT

Connecting Evaluation and Budgeting

No. 30 | March 2014



Marc Robinson

ECD WORKING PAPER SERIES



ECD Working Paper Series No. 30

Connecting Evaluation and Budgeting

By Marc Robinson

www.worldbank.org/ieg/eed

© 2014 Independent Evaluation Group, The World Bank Group

1818 H St., NW

Washington, DC 20433

<http://ieg.worldbankgroup.org>

IEG: Improving the World Bank Group's Development Results Through Excellence in Evaluation

The Independent Evaluation Group is an independent unit within the World Bank Group; it reports directly to the Bank's Board of Executive Directors. IEG assesses what works, and what does not; how a borrower plans to run and maintain a project; and the lasting contribution of the Bank to a country's overall development. The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank's work, and to provide accountability in the achievement of its objectives. It also improves Bank work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

IEG's Evaluation Capacity Development Working Papers are an informal series to disseminate the findings of work in progress to encourage the exchange of ideas about development effectiveness through evaluation.

The findings, interpretations, and conclusions expressed here are those of the author(s) and do not necessarily reflect the views of the Board of Executive Directors of the World Bank or the governments they represent, or IEG management.

IEG cannot guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply on the part of the World Bank any judgment of the legal status of any territory or the endorsement or acceptance of such boundaries.

ISBN-13: 978-1-60244-247-4

ISBN-10: 1-60244-247-9

Contact: IEG Communication, Learning and Strategies

e-mail: ieg@worldbank.org

Telephone: 202-458-4497

Facsimile: 202-522-3125

<http://ieg.worldbankgroup.org>

Acknowledgements

This paper was written by Marc Robinson, a senior consultant at PFM Results Consulting, specializing in public financial management, performance budgeting and fiscal policy. He consults directly to governments, and provides technical assistance on behalf of international organizations including the World Bank, OECD and International Monetary Fund. He has helped governments in more than twenty countries to improve their budgeting systems. Prior to becoming an independent consultant, Dr. Robinson was a staff economist in the fiscal affairs department of the IMF. Before that, he was a senior civil servant and economics professor in Australia. Dr. Robinson has published extensively on budgeting issues. He is a member of the OECD's Advisory Panel on Budgeting and Public Expenditures.

This paper benefitted from review and comments by Nidhi Khattri (Lead Evaluation Officer, IEG) and edits from Heather Dittbrenner (IEG). The task manager for this work was Ximena Fernandez Ordonez (Evaluation Officer, IEG).

None of the views expressed herein by the authors should be construed to represent the policies or positions of the World Bank Group.

Contents

Executive Summary.....	6
1. Evaluation and Performance Budgeting—The Principle.....	8
1.1 Budgeting and Public Sector Performance	8
1.2 Performance Budgeting	9
1.3 Questions Budget Decision Makers Would Like Answered	10
1.4 Evaluations related to Performance Budgeting.....	11
1.5 Limitations of Performance Indicators and the Role of Evaluation	14
2. Evaluation and Performance Budgeting in Practice	15
2.1 The Early Days: Performance Budgeting and Evaluation Closely Linked	15
2.2 Subsequent Performance Budgeting and Evaluation Separation.....	16
2.2.1 France.....	17
2.2.2 United States.....	17
2.2.3 Other Countries.....	18
2.3 The Decline of Evaluation	19
2.4 Neglect of Prioritization in Budgeting.....	20
2.5 The Canadian Experience.....	21
3. Recent Efforts to Connect Evaluation and Budgeting.....	24
3.1 Evaluation and the Budget in Chile	24
3.2 The New Canadian Evaluation Policy.....	26
4. How to Better Connect Evaluation and Budgeting	29
4.1 Making Evaluation More Useful for Budgeting.....	29
4.2 What Type of Evaluation?	30
4.2.1 Outcome Evaluations and the Budget	30
4.2.2 The Role of Process Evaluations	31
4.2.3 Program Logic Analysis and Rapid Evaluation as a Substitute for Outcome Evaluation	32
4.2.4 Evaluation of Program Relevance	33
4.3 The Selection of Evaluation Topics	33
4.4 Ex Ante Evaluation	34
4.5 The Conduct of Evaluations Linked to Budgets.....	35
4.6 Budget-Linked Evaluation and Government-wide Evaluation Policy	36

4.7 Improving the Budget Process	37
4.7.1 A Focus on Priorities	37
4.7.2 Spending Review as a Budgetary Routine.....	37
4.7.3 Program Classification.....	38
5. Conclusions	39
Glossary.....	41
References	43

Executive Summary

Evaluation is an essential tool for good budgeting and a core element of any well-designed government-wide performance budgeting system. Evaluation has a crucial role to play in providing the performance information necessary to maximize the effectiveness of budgeting as a tool for expenditure prioritization and for the promotion of effective and efficient service delivery.

All too often, it is assumed that the only type of performance information necessary to support budgeting is measurement with performance indicators. As invaluable as these are, they have limitations. It is often impossible to assess the effectiveness and efficiency of expenditure based on performance indicators alone. They require careful analysis and interpretation, and this is precisely what evaluation does.

For just this reason, it is inaccurate to view performance budgeting as involving only the use of performance indicators. The correct conception of performance budgeting is as a mechanism for systematically harnessing performance information in general – including evaluation and efficiency reviews *as well as* performance indicators.

Unfortunately, the potential value of evaluation as a budgeting tool has not been realized in practice. In part, this is because evaluation has often not been sufficiently tailored to the needs of budget decision makers. But this also reflects the fact that in many countries, governments have not sought to use the budgeting process to achieve good expenditure prioritization or to promote performance. Thus, limited budget-relevant evaluation reflects the lack of demand for such evaluation by budget decision makers.

Things are changing. The difficult fiscal circumstances facing many countries in the wake of the global financial crisis have created unparalleled pressures for expenditure cuts, not only to support debatable "austerity" policies but also to free resources for high-priority new spending. The pressures for good expenditure prioritization and improved efficiency have greatly increased. It has become vital to connect evaluation and budgeting, to ensure that evaluation fully contributes to the task of cutting ineffective expenditures.

To achieve this, it is necessary that evaluation – insofar as it is specifically intended to support budgeting – change. The choice of evaluation topics and the scope of evaluations need to focus on identifying ineffective or low-priority government programs that should be terminated or scaled back to assist either in reducing government expenditure or in creating additional "fiscal space" for high-priority new expenditures. The differing needs and requirements of evaluation for budgetary purposes, as opposed to evaluation for management and policy improvement purposes, need to be recognized. The notion that a single government-wide evaluation system can serve both of these purposes effectively should be reconsidered.

It is equally important, however, to change budgeting. Only when budget decision makers focus fully on optimizing expenditure prioritization and promoting performance will they appreciate the value of evaluation. Particularly important here is the creation of continuing *spending review* processes designed to identify options for spending cuts. Those responsible for managing this process and advising political

leaders on spending cuts should be the primary budgetary users of evaluation information. The connection between evaluation and budgeting also needs to be sharpened, through the adoption of program classification of the budget, based on outcomes and outputs as opposed to inputs and organizational structures. This is a core element of a government-wide performance budgeting system.

1. Evaluation and Performance Budgeting—The Principle

This section outlines the role which evaluation should, in principle, play in supporting good budgeting. It identifies the key ways performance information in general supports budgeting and then outlines the way performance budgeting seeks to structure the contribution of performance information to budgeting. The section concludes by discussing the nature and role of evaluation as a key component of the performance information base for budgeting.

1.1 Budgeting and Public Sector Performance

Performance information can contribute to good public sector budgeting in three key ways.

First, it helps improve *expenditure prioritization*. Expenditure prioritization means that government resources are allocated to the programs which deliver the greatest benefits to the community given the money spent. This is what economists refer to as "allocative efficiency." For expenditure prioritization, it is the *relevance* and *effectiveness* of programs that are important. Relevance refers to the importance to the community of the outcomes the program aims to achieve – for example, the extent to which to a program addresses a major problem or seeks to achieve outcomes that citizens value highly.¹

Effectiveness, in contrast, means that programs actually achieve their intended outcomes – for example, an AIDS prevention program actually reduces the rate of AIDS infection. Expenditure prioritization, then, will be improved to the extent that budget decision makers have good information on the relevance and effectiveness of expenditure and use this information in deciding which programs and projects to fund.

Second, it puts increased pressure on ministries to improve the *effectiveness* of their programs. If spending ministries know that the Ministry of Finance and other budget decision makers have good information about the effectiveness of their programs and that they take program effectiveness into account when deciding how much funding to provide in the budget, they will feel greater pressure to improve the effectiveness of their programs to protect their budget allocations.

Third, the availability and use of performance information can help ensure that the budget promotes *efficiency*. Efficiency means that government services are delivered at the lowest possible cost, consistent with the maintenance of quality.² Promoting efficiency via the budget means avoiding funding ministries at levels that permit them to operate wastefully (for example, at low levels of labor productivity or with inappropriate business processes, which raise costs unnecessarily). More precisely, budgeting promotes efficiency when budget decision makers seek to base the level of program funding on the efficient cost of service provision while simultaneously holding ministries accountable for maintaining the level and quality of services provided to the public.

¹ For example, a program to tackle high rates of infant mortality has strong relevance, whereas an initiative to promote the appreciation of traditional music will generally rate significantly lower in relevance.

² What we refer to here as "efficiency" is usually described by economists as "technical" or "productive" efficiency.

It might be assumed that good expenditure prioritization and the promotion of effectiveness and efficiency are objectives that would be central to budgeting. Often, however, this is not the case. Traditionally, ministries of finance have not viewed the budget as an instrument for promoting program effectiveness or efficiency. Even expenditure prioritization is, in practice, something that governments are often not good at. A particularly widespread problem is that budget preparation focuses mainly on the consideration of new spending proposals, with little or no review of ongoing (or baseline) expenditure. This imparts inertia to baseline expenditures – often referred to as budgetary "incrementalism" – with the practical consequence that programs that are ineffective or low priority or that have simply outlived their usefulness may continue draining budgetary resources (Robinson 2013a).

The inattention to baseline expenditure does not, however, mean that governments are necessarily good even at scrutinizing new spending proposals to select those that offer the greatest benefits. In some countries, systematic and rigorous processes for analyzing new spending proposals put forward during the budget process are sadly lacking.

Budget decision makers face a range of problems in seeking to improve the prioritization of expenditure and promote effectiveness and efficiency. These problems tend to arise from a number of sources, including political factors and in-built expenditure inflexibilities. However, lack of information tends to be a major part of the problem because, typically, budget decision makers have only limited information on the effectiveness and efficiency of government expenditures. More specifically, they have little idea of the degree of inefficiency (waste) within government ministries or programs, and therefore little idea of the extent to which budget allocations to those ministries or programs could be reduced without adversely affecting the quantity and quality of service provided. They also commonly have limited knowledge of how effective program expenditures are in many areas of government. Addressing this information problem is at the heart of performance budgeting.

1.2 Performance Budgeting

Performance budgeting systematically integrates performance information into budgeting by using it to link funding to results (outcomes and outputs), with the aim of improving performance (Robinson 2007b, 2013c).

Program budgeting is not only the most widespread form of performance budgeting, but it is the form of performance budgeting most applicable to government budgets as a whole. Program budgeting aims to structure the budget in such a way that it facilitates good expenditure prioritization, while simultaneously placing increased pressure on ministries and agencies to improve the effectiveness and efficiency of their expenditures.

The cornerstone of program budgeting is the program classification of the budget, which means the classification of expenditures in the budget primarily according to the public policy objectives and types of services to which funds are directed (Robinson 2013b, 2011). For example, in a program budget the funding provided to the Ministry of the Environment takes the form of specific allocations to a nature

conservation program, an antipollution program, and other similar programs. This results-oriented program classification of expenditure provides a framework for systematically taking information about the performance of programs (or components of programs) into account when deciding budget allocations. So if the performance information makes it clear that some component of a program is ineffective, that should lead to close consideration of whether funding for that component should be terminated. Equally, if the performance information indicates that a program is effective but inefficient, this should lead to a progressive reduction in the level of funding, combined with pressure on the ministry to ensure that it improves efficiency rather than cutting the quantity and quality of program services.

Program budgeting is not the only form of performance budgeting. Other forms include output-based formula funding and purchaser-provider systems – which aim to promote improved efficiency by linking funding to the quantity of outputs delivered using output unit costs.³ In general, these mechanisms are inappropriate for across-the-board application in a government budget and should be used only selectively for appropriate categories of government services. These are forms of performance budgeting for which evaluation has limited potential value (although efficiency analysis is very useful), and which are therefore not discussed in this paper.

However, it should be acknowledged that, in addition to program budgeting, there is another performance budgeting tool that does have government-wide applicability. This is what may be described as *target-based* performance budgeting, which seeks to link the level of funding provided broadly to ministries to centrally imposed targets for the results those ministries are expected to achieve.⁴ In such a system, evaluation has an important potential role in the ex post assessment of performance against targets.

1.3 Questions Budget Decision Makers Would Like Answered

In considering the potential role of evaluation, it helps to identify the specific types of questions to which central budget decision makers would like answers when deciding funding levels for ministries and programs.

With respect to expenditure prioritization and effectiveness, answers to the following questions are of the greatest potential value to budget decision makers:

³ For example, funding provided to public hospitals under a "diagnostic-related group" funding system is determined primarily by the number of treatments of various types the hospitals provide, with different levels of funding for each type of treatment (for example, \$1,000 for treating severe asthma cases and \$5,000 for providing hip replacements).

⁴ For example, funding levels the government agrees to provide for the Ministry of Education for the coming three years might be explicitly linked to targets for improvements in literacy and numeracy levels of school-aged children. The most notable version of this form of performance budgeting is the British Public Service Agreement system as it operated between 1998 and 2007 – the main focus was on outcome targets (such as literacy targets). However, in principle, target-based performance budgeting may also focus primarily on output targets (for example, targets pertaining to the volume of services provided to the public).

- *Which of the spending proposals presented by spending ministries during the budget process are likely to generate outcomes of a sufficient magnitude and importance to justify their costs?*
- *Which existing programs (or components of programs) should be terminated because they are ineffective or not sufficiently effective to justify their cost? 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.*
- *Which existing programs should be considered for closure on the grounds of low relevance? In other words, which existing programs are pursuing outcomes that are not sufficiently important to the community to justify the expenditure concerned?*

With respect to efficiency, what budget decision makers would particularly like to know is what it *should* cost to efficiently deliver specific services (outputs) – at what we will refer to in this paper as the "efficient cost" of service delivery. Expressed differently, decision makers would like to have quantitative estimates of the potential budget savings that could be realized through efficiency improvements (that is, estimates of the difference between the present cost of service delivery and the efficient cost). In addition, they would like to know the time-frame that spending ministries would realistically require to realize these savings and whether any short-term additional funding (for example, funding for implementing cost-reducing information technologies) is needed to achieve these savings.

A program budgeting system will use answers to all the above questions, precisely because it has the broad objectives of improving prioritization, effectiveness, and efficiency. By contrast, because formula funding and purchaser-provider budgeting focus solely on improving efficiency, only information on the efficient cost of service delivery is important. Target-based performance budgeting can in principle make use of information about both the effectiveness and efficiency of expenditure, depending on the extent to which it focuses on outcome targets (in which case effectiveness information is most important) or output targets (in which case efficiency information is more important).

1.4 Why Evaluation is important for Performance Budgeting

What should the role of evaluation be in meeting the information needs of performance budgeting systems? Answering this requires clarity about what evaluation is and how it relates to other forms of expenditure analysis.

Evaluation is usually defined in a broad sense, exemplified by the widely used 2000 definition by the Organisation for Economic Co-operation and Development (OECD) that defines evaluation as the "process of determining the worth or significance of an activity, policy, or program (which is) as systematic and objective as possible." The evaluation literature typically identifies three main categories

of evaluation – outcome evaluations, process evaluations, and ex ante evaluations – each of which is potentially relevant to budgeting.

Outcome evaluations: Also known as "impact" evaluations⁵ and "summative" evaluations, outcome evaluations aim to ascertain the extent to which a program or project's intended outcomes have been achieved – that is, the extent to which the program or project has in fact brought about desired changes.⁶ Cost-benefit analysis and cost-effectiveness analysis are forms of outcome evaluation.

Process evaluations: Also known as "formative" or "implementation" evaluations, these are usually analyses of program or project implementations that are aimed at improving performance. The primary focus of process evaluations is the identification of policy design or management changes that would improve the effectiveness of programs. However, process evaluations frequently also address efficiency issues.

Ex ante evaluations: Also known as "prospective" evaluations and sometimes simply as "appraisals," these evaluations are carried out before the program or project concerned has actually been implemented.

One other category of evaluation should be added – *program logic analysis*, which is also known as "theory-based evaluation" (World Bank 2004: 10), "intervention logic analysis" or sometimes as "design evaluation." Program logic analysis looks at the manner in which a particular program or project is supposed to achieve its intended outcome and asks whether, in the light of relevant theory and experience, it is likely that it will achieve those outcomes. The starting point of program logic analysis is the explicit description of the relevant program's "program logic" – that is, the causal links by which program outputs are expected to generate intermediate program outcomes and, through that, higher-level outcomes. Once the program logic is made explicit, the plausibility of the presumed links between outputs and outcomes, and between intermediate outcomes and higher level outcomes, is assessed by reference to relevant theory.

Program Logic Analysis: Two Examples

Example 1: Unemployment Rates

⁵ In this paper, no distinction is made between impacts and outcomes, and impacts are considered to be simply a type of outcome. This reflects the fact that this distinction is rarely made in the performance budgeting literature and that it is a distinction that can be hard to make at the margin. Consistent with this – and to keep things as simple as possible – no distinction is made here between outcome and impact evaluations. It should, however, be noted that this is a distinction that is frequently made in the evaluation literature.

⁶ For example, an outcome evaluation of a program designed to raise the educational levels of disadvantaged children might examine the levels of literacy and numeracy levels achieved by children targeted in the program and seek to determine – perhaps by comparisons with similar children who have not benefited from the program – the extent to which any improvements in their literacy and numeracy levels can reasonably be attributed to the program. Additionally, this paper does not distinguish between long term effects and attribution analysis, or specific methods.

Consider a labor market program that has the explicit objective of lowering the rate of unemployment by providing wage subsidies for the long-term unemployed. In this case, a program logic analysis would ask – primarily on the basis of labor market economics and experience in other countries – whether such a program is likely to achieve this objective or alternatively whether it is instead likely only to achieve a "churning" of employment (with subsidized workers replacing nonsubsidized workers). Such an analysis is distinct from directly assessing the outcomes of the program by, for example, comparing changes in the net employment levels of firms that hire subsidized workers with changes in employment levels in terms that do not (which would constitute a form of outcome evaluation, rather than program logic analysis).

Example 2: Childhood Obesity

In examining a campaign of TV advertisements that aim to reduce child obesity by, in essence, embarrassing the parents of overweight children, a program logic analysis would consider whether relevant psychological and marketing theory suggests that the campaign will succeed in changing the behavior of parents and, through this, in reducing the number of overweight children.

Program logic analysis overlaps with the concept of ex ante evaluation, because the latter should always include an analysis of program logic. However, program logic analysis is also applied to existing programs.

This list does not include one form of expenditure analysis that is enormously important for budget: *efficiency reviews*. These reviews aim to identify a means of reducing the cost of delivering outputs (the goods or service delivered by government) or carrying out a process without reducing its quality – for example, by investing in cost-reducing technology, eliminating unnecessary processes (for example, through "business process re-engineering"), or tightening management or incentives to ensure that low-productivity staff work properly. A pure efficiency review has nothing whatsoever to do with effectiveness. Its focus is solely on the cost of the output, with no attention to the extent to which the output is delivering its intended outcome.

Are efficiency reviews also a type of evaluation? To debate this question would require the setting of precise but entirely artificial boundaries for evaluation. On the one hand, in the evaluation literature, the assessment of efficiency is regarded as a legitimate potential element of process evaluations. On the other hand, analysis that is labeled evaluation tends to be primarily focused on outcomes/effectiveness issues – as is implicit in definitions of evaluation that focus on the assessment of "worth or significance." Moreover, most efficiency analysis is in practice not called "evaluation" and is not carried out by persons who call themselves "evaluators."

Reflecting this, the term evaluation as used in this paper does not include efficiency analysis in the sense of analytic studies focused exclusively on efficiency issues. Evaluation, rather, focuses primarily on

outcomes and effectiveness issues and usually takes one the forms described above.⁷ "Evaluation" and efficiency analysis are thus treated essentially as separate forms of expenditure analysis here, overlapping only to the extent that process evaluations – though primarily focused on outcomes/effectiveness issues – may include some elements of efficiency analysis. The corollary of this is that good budgeting should be considered to require *both* evaluation and efficiency analysis.

1.5 Limitations of Performance Indicators and the Role of Evaluation

Evaluation is important for budgeting precisely because, taken alone, performance indicators are generally insufficient to establish the effectiveness and efficiency of government programs.

In particular, outcome performance indicators either do not distinguish or distinguish only to a limited degree between the outcomes achieved by a government program and the influence of external factors. External factors are external events or client characteristics that influence the measured outcomes of programs but are beyond the control of government. For example, the success of labor market programs in reducing rates of long-term unemployment will be greatly influenced by the state of the economy; such a program is likely to have very little measured success during a period of recession. The scope for adjusting outcome indicators to eliminate the effect of such external factors is, unfortunately, quite limited. It is precisely the task of outcome evaluations to analyze imperfect outcome indicators in order to distinguish outcomes from the intervention versus the impact of external factors. In short, assessing the effectiveness of existing programs is not something that can—except in rare cases—be achieved using outcome indicators alone. Interpretation is essential, and the best possible interpretation comes from the application of evaluation techniques.

To make this point is not in any way to downplay the crucial role of performance indicators – especially outcome indicators – in performance budgeting. Although it is true that indicators alone are insufficient to judge program effectiveness, it is also true that outcome evaluation depends completely on the availability of good outcome indicators. Performance indicators and evaluation are therefore not alternatives, but inseparable parts of the performance information base.

As noted, from a budgetary point of view the task is not only to identify ineffective programs, but also to determine which programs can be fixed and which cannot. Indicators alone are insufficient to judge whether an ineffective program can be fixed through policy or management changes. Only through careful evaluation of the program's processes can such an assessment be made.

⁷ This is not to suggest that evaluation should be thought of as involving only the application of analytic techniques that are formally labeled "evaluation" methodologies, nor that it is an activity that is carried out exclusively by professionals who are trained as or consider themselves to be evaluators. For example, cost-benefit analysis and cost-effectiveness analysis are techniques predominantly employed by economists. Similarly, program logic analysis – or analysis very similar to what is defined here as program logic analysis – is often carried out by professionals who regard themselves as either policy analysts or as specialists in economic policy, social policy, or some other discipline.

The same is true in respect to the importance of efficiency review for budgeting. Efficiency indicators – such as unit costs or labor productivity measures – are necessary but not sufficient as a means of assessing the efficiency of government programs and processes. For example, a time series of unit cost measures showing that the unit cost of a specific government service has progressively increased over recent years does not, in itself, provide conclusive evidence of deteriorating efficiency. Although this is one possibility, there are other potential explanations for such a trend – such as changes in case complexity or quality, or input cost changes. In almost all cases it will be essential to analyze the cost data and examine the underlying work processes – in other words, to carry out an efficiency analysis – to establish with confidence what the real problem is.

Performance indicators can be seen as raising questions about effectiveness and efficiency rather than as providing answers to those questions. It is, in general, only through analysis – and in particular through evaluation and efficiency review – that these questions can be answered.

This section has suggested that *good performance information* is crucial to good budgeting. Performance budgeting seeks to structure and systematize the contribution of performance information to budgeting. Evaluation – as a specific form of performance information – has a particularly important role to play in government-wide performance budgeting.

2. Evaluation and Performance Budgeting in Practice

This section reviews the actual relationship between evaluation and performance budgeting by looking at the experiences of countries that have made substantial efforts to implement both. As discussed, evaluation is *in principle* essential to performance budgeting. But what is the position *in practice*? It has been almost 50 years since both performance budgeting and evaluation made their entrances into the field of public administration. To what extent has evaluation in fact improved the quality of expenditure prioritization and helped make budgeting an important tool for promoting effectiveness and efficiency? And to the extent that evaluation may have failed to live up to its potential as a tool for supporting budgeting, why has this been the case, and what lessons does this provide for the future?

2.1 The Early Days: Performance Budgeting and Evaluation Closely Linked

Performance budgeting and evaluation were linked at birth. Although there were earlier forms of performance budgeting, it was the arrival in the United States in the late 1960s of program budgeting in the form of the planning, programming, and budgeting systems (PPBS) that marked the true debut of performance budgeting on the international stage. Over the subsequent decade, governments across the world followed the U.S. example, introducing program budgeting systems broadly based on the PPBS model.

PPBS from the outset strongly emphasized the role of evaluation, and it was the arrival of PBBS that triggered the creation of the "first generation" of government evaluation systems internationally.

In the United States, program evaluations, including benefit/cost analysis, were from the start integral elements of the PPBS system. Although it was intended that performance indicators would play an important role, no one considered that indicators alone would provide the information needed. Rather, it was thought that it would be primarily through evaluation that government would be enabled to make more rational decisions about expenditure prioritization than had been possible (Steiner 1967). It was as a direct result of PPBS that the evaluation and policy analysis function grew rapidly, for the first time, within U.S. government agencies. At least 800 analysts and evaluators were appointed in 16 U.S. domestic agencies during the first 5 years of the PPBS system (Marvin and Rouse 1970). By the late 1970s, the U.S. government was spending around \$200 million per year on evaluations (Wholey 1978: 52).

Something similar occurred in many other countries which followed the US lead in developing program budgeting systems at that time. Canada was the second country to introduce a program budgeting system (also called PPBS) in 1966 (Good 2008: 252), immediately following this in 1969-70 with the first steps toward a formal evaluation system (Lahey 2010; Auditor General of Canada 2009: 4) in which "program evaluation was meant to be a major source of evidence on program effectiveness in the context of decision-making on priorities and expenditures" (Aucoin 2005). Another early mover was France, which adopted program budgeting in 1968 under the name *la rationalisation des choix budgétaires*. This program, which survived until the mid-1980s, emphasized evaluation, with particular emphasis on improved ex ante evaluation (Huet 1971).

Although the main influence of the PPBS model was felt in the 1960s and 1970s, the close partnering of program budgeting and evaluation could be seen even in PPBS-style models introduced later, such as in Australia in the second half of the 1980s. Like its earlier counterparts, the Australian program budgeting system was also integrally linked to the country's first government-wide centrally driven evaluation system. Developed progressively from 1987, this system required ministries to systematically evaluate all existing spending programs. Moreover, it was mandated that all new budgetary spending proposals include a statement of objectives and performance measures, as well as proposed arrangements for their future evaluation (Department of Finance and Australian Public Service Board 1987; MacKay 1998, 2004).

2.2 Subsequent Performance Budgeting and Evaluation Separation

Since those early days, performance budgeting and evaluation have drifted apart. Only in a few countries has evaluation played a substantial part in the preparation of the budget, or in funding decisions more generally. Indeed, some contemporary performance budgeting systems place little or no emphasis on the role of evaluation, but instead see the systematic input of performance information

into the budget as coming essentially from performance indicators. It is not uncommon to find performance budgeting (erroneously) defined as the use of performance *indicators* – rather than performance information more generally – in budgeting.

2.2.1 France

A striking example of the neglect of evaluation in some contemporary performance budgeting is found in the French performance budgeting system that came into full force in 2006. The new system, usually referred to as the LOLF system (after the 2001 legislation on which it is based), is essentially founded on a program classification of expenditure and – importantly – a dramatic reduction of traditional "line item" budget controls (Chevauchez 2007; Lannaud 2007). During its implementation over the five-year period 2001-2006, LOLF assigned essentially no role to evaluation. Instead, emphasis was placed exclusively on program performance indicators (including, for example, in the annual program performance reports, which were the centerpieces of the system). During the period between the first failed attempt to introduce performance budgeting in the form of the RCB mentioned above and the introduction of the LOLF system, evaluation had withered away in France (see below), and the implementation of LOLF over 2001-2005 did nothing to change this situation. Only as an afterthought was the LOLF modified in 2009 to introduce a requirement for *ex ante* evaluations of all new spending proposals presented to parliament. As useful as this step was, it left unchanged the neglect of evaluation of baseline expenditure.

In 2011, 10 years after the passage of the LOLF legislation, it looked as if France would establish a system for evaluation of baseline expenditure. In 2011, the Sarkozy government's ad hoc spending review – *révision générale des politiques publiques* (RGPP) – foreshadowed certain steps to rebuild evaluation, including the creation of an inter-ministerial evaluation committee and the assignment of the responsibility of developing the system to a specific unit within the directorate-general of modernization of the state under the finance minister (RGPP 2011: 15-16). This was an explicit recognition that evaluation could and should play an important role in facilitating the reprioritization of expenditure during a spending review process such as RGPP. It also recognized that, in practice, evaluation was unable to play that role because France's system was incapable of delivering relevant evaluations to budget decision makers. In addition, the ad hoc and short-lived nature of the RGPP process meant that there was insufficient time to commission substantial evaluation activity to support its work.

In practice, little action was taken during the remaining period of Sarkozy's presidency to develop the integration of evaluation. Since that time, the government of President Hollande has placed some emphasis on the role of evaluation as part of the mechanism of *Modernisation de l'Action Publique*, which it introduced to replace the RGPP. At the time of this writing, it is too early to assess these developments.

2.2.2 United States

In the United States, budgeting and evaluation also drifted apart, and the role of evaluation in U.S. public management had by 2000 diminished greatly compared to the 1960s and 1970s. Thus in 2009, the head of the Office of Management and Budget (OMB) bemoaned the weak role of evaluation in the budget in the following terms:

... Evaluations can help the administration determine how to spend taxpayer dollars effectively and efficiently – investing more in what works and less in what does not. ...[However] many important programs have never been formally evaluated – and the evaluations that have been done have not sufficiently shaped federal budget priorities or agency management practices... As a consequence, some programs have persisted year after year without adequate evidence that they work (Orszag 2009).

This was notwithstanding the fact that the Bush Administration's performance budgeting initiative – the Program Assessment Rating Tool (PART), introduced in the early 2000s – had recognized evaluations as one element of the flow of performance information that should inform the budget process. Under PART, all federal programs were given a performance rating on a four-point scale, ranging from "ineffective" to "effective," and OMB assessors were required to take into account any program evaluation, together with relevant performance indicators and other information, in determining program ratings (OMB 2002).⁸ The PART rating was essentially a summary program indicator intended to clarify program performance problems to budget decision makers, including the Congress (which in the United States has great independent budgetary power). The intention was that PART would tangibly link budgeting and performance by ensuring that the continued budget funding of programs that received poor performance ratings would be closely scrutinized.

Despite its recognition of evaluation as one element of the performance information base, PART did not require agencies to carry out more or different evaluations. It was primarily a performance measurement, rather than an evaluation, initiative. The choice of evaluation subjects and the focus of evaluations remained entirely a matter for spending agencies to determine – as in fact did the initial decision as to whether to conduct evaluations. OMB played no role in initiating or managing evaluations to serve as inputs to expenditure prioritization decisions.

2.2.3 Other Countries

This neglect of evaluation in performance budgeting was also true of other – rather different – performance budgeting systems introduced internationally in recent decades. For instance, the U.K. public service agreement system, which operated between 1997 and 2007, neglected it, focusing on budget-link performance targets (Smith 2007). Indeed, it could be said that internationally, there has been considerably more focus on, and use of, efficiency analysis by ministries of finance over the past

⁸ In addition, although it was more a performance management and performance budgeting initiative, the Government Performance and Results Act of 1993 had given some encouragement to evaluation by requiring that program performance reports include a "summary findings of those program evaluations completed during the fiscal year covered by the report."

two decades than there has been of evaluations. This is obviously true of performance budgeting systems that are primarily focused on efficiency enhancement, such as the purchaser-provider system that Australia unsuccessfully attempted to implement from the late 1990s to early 2000s (Robinson 2007c). It has also been true elsewhere, including in the United Kingdom, where the Treasury established wide-ranging efficiency reviews in the mid-2000s.

2.3 The Decline of Evaluation

This relative neglect of the use of evaluation in budgeting today marks a huge change from the days of the brave new world of budget-linked program evaluation systems in the 1960s and 1970s. What happened along the way? In many countries, government-wide evaluation systems were either dismantled or scaled back and sidelined. This was notably the case in the United States: By the end of that decade, the program evaluation divisions that had existed in every government agency had "all but vanished" (Weinstock 2003), and one contemporary commentator (Wye 1989) observed that "budget cutbacks have reduced evaluation resources and internal constraints have diminished evaluation utilization" to such a point "that it is not too early to be concerned about the future of evaluation in the federal government."

In France, the story was a little less dramatic but fundamentally similar, with the evaluation effort gradually fading away notwithstanding a couple of half-hearted and unsuccessful attempts (in 1990 and 1998⁹) at resuscitation. A decreasing number of evaluations were carried out. No longer was evaluation presented as a tool for budgeting. Instead, the emphasis was solely on its use in management and policy improvement, and even in that domain it appeared to be little used. By the early 2000s – when the new LOLF performance budgeting system was introduced – the evaluation system in France existed only on paper. Symptomatic of this decline was the fact that bureaucratic responsibility for the residual central evaluation function was assigned to the *Commissariat Général du Plan* – a body that had in the early post-war decades been an all-powerful economic planning body but that had by the early 2000s become a marginal player in the French bureaucracy (it was abolished in 2006).¹⁰

This retreat from evaluation partly reflected a "growing disillusionment with social science evaluation" in government (Floden and Weiner 1982: 367). Early hopes that scientific outcome evaluation would yield clear and robust conclusions about program effectiveness and cost-effectiveness were disappointed as it became clear just how difficult it was, methodologically, to handle uncertainty and to factor in intangible costs and benefits (Toulemonde and Rochaix 1994: 49).

⁹ In 1990, as part of the Rocard government's public service renewal reforms, there was a government decree mandating certain formal reforms to the evaluation system – including the creation of an interministerial evaluation entity, the *Conseil Scientifique de l'Evaluation* – in an attempt to ensure its relevance. In 1998, a further decree, together with prime ministerial circular, made a further attempt to salvage the evaluation system. It is important to note that neither of these attempts linked evaluation explicitly to the budget process.

¹⁰ In more formal terms, the *Commissariat* was "replaced" by *Centre d'Analyse Stratégique* (essentially a research body reporting to the prime minister), which virtually dropped the evaluation function.

Why this disillusionment? Part of the problem was the limited availability of outcome indicators that could serve as the basis for outcome evaluations. (Although this a problem continues, it was vastly more serious at that time, when government performance indicator systems were largely nonexistent.) Another reason for disillusionment was that evaluation was widely viewed as lacking a client orientation. In the United States, for example, evaluation documents produced under PPBS came to be viewed more as a mass of indigestible paperwork that obstructed good management than as a useful information source for decision makers. The evaluation discipline came to be seen as insufficiently focused on providing information useful to managers and budgeters, and more focused on meeting its own internally defined standards of what constituted good evaluation work. The resulting emphasis on thorough, “scientific” evaluation reports meant that the reports were too cautious to reach usable conclusions, “too costly and time-consuming compared to their real use and effect” (OECD 1998: 3), and often took so long to produce that they lost their timeliness. With respect to timeliness, the judgment made in the French government's 1998 evaluation circular (Gouvernement de France 1998) was representative:

“One of the big problems of the evaluation system as it has operated up to this point has been the excessive duration of the process of evaluation. The accumulated delays at various stages of the process have on average resulted in a delay of three and a half years between the preparation of the terms of reference of an evaluation and the publication of its findings. Under these circumstances, it has only rarely been possible to use evaluation findings in the decision-making process”.

2.4 Neglect of Prioritization in Budgeting

The problem did not, however, lie exclusively with evaluation; there were also significant failures of budgeting systems to focus sufficiently on the objective of improving expenditure prioritization. Notwithstanding affirmations of the supposed importance of expenditure prioritization, in practice governments often preferred to avoid hard choices between programs. When adjustments had to be made, small across-the-board reductions applied to all ministries and programs were politically easier than identifying and cutting low-priority and ineffective programs. Alternatively, governments advanced the proposition that any necessary savings could be made entirely through improved efficiency and focused on cuts to specific categories of inputs – for example, an across-the-board cut of several percent to spending ministry's non-personnel expenditures. The problem was aggravated by the weak aggregate expenditure discipline that prevailed in many OECD countries and that obscured the importance of cutting ineffective or low-priority expenditure to finance important new priorities.

Evaluation was only going to be routinely used to improve expenditure prioritization if the budget preparation process itself emphasized the reallocation of expenditure from low-priority and ineffective programs to more socially important expenditures. When this did not happen, the failure to link evaluation to budgeting was necessarily a demand as well as a supply problem.

2.5 The Canadian Experience

The story of evaluation and its relation to budgeting in Canada is particularly interesting, for several reasons. In the first place, Canada maintained a formal government-wide evaluation system from the late 1960s to the present time. Second, there has been both a recurrent awareness of the failure of evaluation and budgeting to connect in the way originally intended and a number of efforts to address the problem. Indeed, the most substantial of these efforts is under way at present, with progressive implementation of reforms to the evaluation system mandated in 2009, a central aim of which is to ensure that evaluation properly supports the new Canadian expenditure management system. (These developments are discussed in Section 3).

As elsewhere, disappointment with the failure of the first evaluation system to deliver its intended benefits was evident in Canada by the 1980s. In 1993 a review by the Auditor General found that "the story of program evaluation in the government of Canada is one of high expectations and great potential that have been only partly fulfilled" (Auditor General of Canada 1993). The review found that—

- Program evaluations were "frequently [...] not timely or relevant"
- Evaluations tended to focus on minor activities, and not on the programs that accounted for most government expenditure
- Evaluations had a predominantly operational (that is, management/policy improvement) focus
- Far too little emphasis was given to the evaluation of program effectiveness
- Evaluation was largely decentralized, with both the selection of evaluation topics and the focus of evaluations decided overwhelmingly by spending ministries rather than by the ministry of finance¹¹ or other central agencies
- The fact that evaluations were controlled at the ministry level gave rise to perceptions that their findings were often "dampened down."

Evaluation was widely perceived to have failed to substantially improve expenditure prioritization. As early as 1984, a government task force on program review found that the lack of focus on effectiveness in program evaluations greatly limited their usefulness to central decision makers (see Aucoin 2005). The 1993 Auditor General report particularly lamented the failure of evaluation to contribute to the "management of government expenditure" at a time when public finances were on an unsustainable trajectory and there was a wide recognition of the need for "informed decisions aimed at controlling growth of the public debt."¹²

¹¹ In Canada, this means in particular the Treasury Board and the Treasury Board Secretariat.

¹² This was in spite of 1991 revisions to the government's evaluation policy that had supposedly aimed to make evaluation more strategic and useful to central decision makers. One element of this revised policy was a new provision for the conduct of centrally-mandated evaluations – a provision that remained largely a dead letter.

What was meant by the lack of focus on effectiveness was that relatively few outcome evaluations – evaluations that aim to ascertain the extent to which intended outcomes have been achieved – were being conducted. Instead, the primary focus was on process evaluations, which recommended methods of improving program effectiveness but did not actually directly assess effectiveness. This was, however, not accidental. In a system where control of evaluation was decentralized to the spending ministries, it was natural that evaluations would reflect the interests and preoccupations of these ministries. Spending ministries are naturally more focused on improving their programs than on analysis which questions the existence of these programs. Moreover, the ability to carry out outcome evaluations was severely undermined by the lack of outcome data, reflecting the general weakness of the government indicator system and the tendency to focus more on output and process indicators than on outcome indicators. Thus a 2009 report by the Auditor General found that most evaluations were "limited in their assessment of program success and effectiveness" precisely because they were "hampered by inadequate data." This echoed the findings of a 2005 review by the Treasury Board Secretariat that concluded that few programs had reliable systems for collecting effectiveness data and that this was severely undermining the quality of evaluation reports (Auditor General of Canada 2009: 11-12, 33).

The limited budgetary relevance of evaluation was graphically illustrated in 1994-1996, when Canada undertook deep expenditure cuts to restore its public finances in what was known as the program review process (Bourgon 2009; Good 2008: 266-270). Systematic evaluation studies played a negligible role in this process. This was principally because, with decisions on expenditure cuts being made very rapidly, there was no time to commission evaluations to inform the process. At the same time, the existing stock of program evaluations was of very limited value in the process. The fact that evaluation was found wanting at a time when the budget really needed to make use of it was consistent with experience during a similar, smaller-scale, fiscal consolidation experience 10 years before.¹³

The problem lay partly with the nature of evaluation but could not be blamed exclusively on the evaluation system. As a 2004 government review noted, "part of the problem lies in the lack of demand for effectiveness information" (Treasury Board of Canada Secretariat 2004: 2). In other words, part of the reason effectiveness evaluations were not carried out for budget use was that budget decision makers were not using effectiveness as a key criteria for expenditure decisions.

The growing recognition of these difficulties has led Canada in recent times toward a significant effort to reconnect budgeting and evaluation.

¹³ The 1993 Auditor General report (paragraphs 8.61-8.62) noted that "in 1985 the Ministerial Task Force on Program Review referred to program evaluation as a source of information to make extensive recommendations on government programs.... Among the areas where it sought advice were cases of duplication among programs, programs that might be eliminated, and programs whose scope could be reduced. The Task Force indicated that it had made significant attempts to assess programs based on the findings in program evaluation reports from departments. However, as a user of program evaluation information, the Task Force found the material provided to them by government evaluators did not satisfy their needs."

In summary, international experience in connecting evaluation and budgeting has been disappointing, with evaluation generally making a limited contribution to expenditure priority decisions. The reasons for this have been the following:

- Neither the choice of evaluation topics nor the scope of evaluations has been directed to serve the needs of budget decision makers. The main focus has been on process evaluations that focus on identifying management policy changes to improve program effectiveness, rather than on outcome evaluations that assess the degree of program effectiveness. In addition, evaluations have often not been targeted at programs with the greatest budgetary significance.
- The lack of focus on outcome evaluations, and the disappointing quality of many outcome evaluations, has been in caused partly by the relative lack of good outcome indicators.
- Ministries of Finance have tended to have little involvement in the selection of evaluation topics or in determining the scope of evaluations.
- Evaluation has been primarily a decentralized process, under the control of spending ministries. This has been a major reason for the tendency to focus on process evaluations, because these are useful for internal management/policy improvement purposes and at the same time are not as potentially embarrassing as outcome evaluations.
- The control of evaluations by spending ministries – in other words, by the entities being evaluated – has at times led to the perception that evaluations are insufficiently critical and objective.
- The problem has also been that budgeting process have often not been strongly focused on improving expenditure prioritization. This means that effectiveness has often not been a key criteria for budgetary expenditure decisions.
- When governments have implemented major spending cuts, they have often done so via ad hoc review processes that have made little use of evaluation. This has been in part because the lead times for carrying out targeted evaluation has been too long to permit the use in spending review processes, which have typically been of short duration (for example, one or two years).

This points to a paradox that must be faced in considering the role of evaluation performance budgeting. On the one hand, it is unambiguously clear at the theoretical level that evaluation is essential to performance budgeting and that performance indicators can never be sufficient. On the other hand, at a practical level there has been a widespread failure to connect evaluation to budgeting and to integrate it with performance budgeting systems. In seeking to resolve this problem, it is obviously essential to address the challenges highlighted by past experience.

3. Recent Efforts to Connect Evaluation and Budgeting

In the wake of the global financial crisis, renewed interest in the role of evaluation as a budgetary tool is increasingly apparent among OECD countries, with countries such as the United States and the Netherlands making moves in this direction. However, two countries stand out for their sustained efforts over a longer period to ensure that evaluation contributes substantially to budgeting. One is Chile, where the sustained effort in this direction has been under way for 15 years. The other is Canada, which in 2009 initiated a new evaluation policy designed to address the problems of the past and to reconnect evaluation and budgeting; the country had made important moves in this direction even before 2009. The approaches taken by these two countries point to some important lessons for other countries, while at the same time raising some significant questions.

3.1 Evaluation and the Budget in Chile¹⁴

In Chile, since 1997 the government has strongly emphasized evaluation as a tool for "decision making in the allocation of public resources" as well as for management and policy making (Guzmán 2007). Since then, the Ministry of Finance has developed and managed a government-wide evaluation policy aimed at serving these objectives. Evaluation is seen as part of a broader performance management and budgeting system known as the "system of evaluation and management control," the declared aims of which are to improve the effectiveness of policy making and management throughout central government, to create performance incentives for civil servants, and to make the budget results oriented (Arenas and Berner 2010).

Under this system, the Ministry of Finance manages evaluations and takes the lead in identifying the programs to be evaluated each year. There are three different types of ex post evaluation in the Chilean performance management system:

- *Impact (that is, Outcome) Evaluations:* The main focus of these is whether programs have achieved their intended outcomes. However, the terms of reference of these evaluations routinely require evaluators also to make recommendations for program improvement. They therefore represent combined outcome and process evaluations.
- *Evaluations of Government Programs:* These are essentially "rapid evaluations," the core element of which is program logic analysis, although they sometimes examine efficiency issues.
- *Institutional Evaluations:* Called "comprehensive expenditure evaluations these are essentially process evaluations that look at specific institutions or sectoral groups of institutions. They examine a range of issues including the consistency of institutional and sectoral objectives, organizational structures, production and management processes, resource use, and service delivery performance.

¹⁴ This section draws on the author's participation in an OECD mission to Chile in April 2011 (Hawkesworth, Huerta Melchor, and Robinson 2012).

In 2009 the Ministry of Finance introduced a new formal mechanism for ex ante evaluation of new spending proposals and subsequently added a technical assistance service to entities under which the ministry provides advice on how to develop and present good quality new spending proposals. A key feature of this system is that the types of evaluation carried out--and the fact that the ministry writes the terms of reference -- ensures that effectiveness is a strong focus, in an effort to safeguard against the drift into primarily process evaluations that has been a trend elsewhere.

The number of evaluation has progressively increased over time, with 33 in 2009 and 39 in 2010. The evaluated programs and organizations are selected by a designated interministerial committee consisting of the Ministry of Finance and other central agencies; the proposed list is then the subject of consultation with the Congress. External evaluators (consultants or research institutions) conduct the evaluations under contracts with the Ministry of Finance, which provides terms of reference and methodological guidelines. The Ministry of Finance and the relevant spending ministry discuss the evaluation recommendations and agree on the improvement actions to be taken in response to the recommendations. This then becomes the subject of a formal agreement, the implementation of which the Ministry of Finance monitors.

A key element of the Chilean system is the imposition of tight time constraints on evaluations to avoid the problem of evaluation reports being finalized only years after they were commissioned and losing relevance as a result. Most evaluations in Chile are finalized within 4-10 months. This timeliness is due in part to the quite large role played by the so-called "evaluations of government programs" (see above), which are conducted primarily as desk reviews, without the need for additional data collection.

Although a key aim of this system is to ensure that evaluations serve the needs of the budget, the system developed by the Ministry of Finance is not purely or even primarily intended to serve the information requirements of budget decision makers. Rather, as indicated, its aims have been to develop evaluation as a tool for management, policy, and budgeting functions generally. When the system was first developed, government institutions were undertaking relatively little evaluation for any purpose. In this context, the aim of the Ministry of Finance was not just to meet the performance information requirements of budgeting, but also to promote the development of evaluation more broadly within Chilean government.

Despite the fact that evaluation in Chile is intended to serve both budgetary and management improvement objectives, there is a perception that it is not having a sufficient impact on the budget. It appears that although evaluations quite often do generate program design and management changes, they have much less impact on funding of programs. A study of evaluations conducted between 2000 and 2009 found that only 7 percent led to the termination or replacement of a program. By contrast, 37 percent led to design or process modifications, 25 percent to "substantial" program redesign, 24 percent to "minor" changes, and 7 percent to institutional reassignment of the program (Arenas and Berner 2010: 69M). In addition, the Ministry of Finance's budget analysts do not often discuss evaluation findings with the affected institutions during the negotiations with line ministries.

A key reason for this is that the Chilean evaluation system has evolved in such a way that it is focused more on policy and management improvement. This can be seen, first, in the choice of programs to be evaluated. Evaluations are not targeted at programs that are regarded as potential candidates for budgetary savings. Rather, the choice of topics appears to reflect a goal of evaluating most programs over time, combined with a desire to pay greater attention to programs that are seen as potentially needing policy redesign or management improvement. There has, in addition, been a deliberate policy of including within the annual schedule of evaluations some programs that are seen to be good performers.

In addition, and notwithstanding the steps that have been taken to ensure that effectiveness is one of the key focuses of evaluations, the primary focus has been on evaluation as a source of recommendations to improve management. Not only the outcome evaluations, but the other key categories of evaluation conducted under the system, such as the institutional evaluations, have broad objectives, with a heavy emphasis on management improvements. Moreover, evaluations appear to be more focused on rescuing troubled programs than on advising whether it is appropriate to rescue them. As a result, evaluation reports are seen by budget staff as having limited assistance in the preparation of the budget.

One of the questions raised by these aspects of the Chilean experience is whether it is desirable to combine the system of budget-oriented evaluation with a more general government-wide evaluation system.

Another problem that has affected the budgetary use of evaluation in Chile is the lack of alignment between evaluations and the programs regarding which expenditure is classified in the budget. The programs that are evaluated are not only different from budgetary programs, but are in significant instances spread across several budgetary programs. As a result, even Ministry of Finance budget analysts are not always clear to which budgetary program certain evaluation programs correspond—so sometimes even ministry staff have trouble knowing how to give budgetary effect to the evaluations when they conclude that evaluation programs are ineffective or that their program logic is dubious. And to outsiders – including in the president's office and the Congress – the relationship between evaluation findings and the budget is even more opaque.

For these reasons, there is a sense in Chile today that the evaluation system could and should be made much more useful for budgeting. The government and the ministry of finance are considering what type of changes should be made to achieve this objective.

3.2 The New Canadian Evaluation Policy

The efforts of Canada to reconnect evaluation and budgeting are more recent and have grown out of the mounting frustration from the 1990s over the disappointing contribution of evaluation to the budget process, as noted above. This led, in the early 2000s, to a series of measures designed to make evaluation more useful for budgeters. These culminated in the issuance of a new evaluation policy in

2009 by the Treasury Board Secretariat – a Canadian institution that is essentially responsible for the policy component of the ministry of finance function.

Even prior to the adoption of the new policy in 2009, a number of measures were taken to increase the budgetary relevance of evaluation: the imposition in 2000 of a requirement that ministries evaluate the effectiveness of all transfer payments; and a 2001 directive that all evaluations henceforth consider the relevance, success, and cost-effectiveness of programs. Both these measures explicitly sought to address the problem of evaluations being overwhelmingly process focused rather than effectiveness focused. In a similar vein, mandatory ex ante evaluations of the effectiveness of selected new federal regulations were introduced in 2007 (Auditor General of Canada 2009: 7).

The new evaluation policy issued in 2009 represented, however, a much more across-the-board attempt to overhaul the role of evaluation and raise its contribution to budgeting. The central aim of the new policy is that evaluation systematically supports the new Expenditure Management System introduced in 2007. In particular, it was intended that evaluation would strongly support the new system of strategic spending reviews and would more generally "inform government decisions on resource allocation and reallocation." The requirement that all evaluations cover effectiveness as well as process issues was strongly reinforced by the new policy (Treasury Board of Canada Secretariat 2009; Lahey 2010).

The 2009 policy aims for comprehensive coverage of program evaluations. It is now mandatory that all direct program expenditure – and not merely transfer expenditure – be evaluated. Ministries are required to carry out these evaluations over a five-year cycle. This approach of mandatory comprehensive evaluation over a multiyear cycle is similar to the approach implemented in Australia in the 1980s. In the Canadian context, it represents a reversion to the objective of comprehensive evaluation coverage that had been set in the 1970s but that had in the 1990s been replaced with a strategy of selective evaluation (Auditor General of Canada 1993).

The aim of comprehensive coverage of evaluation highlights the fact that, although strengthening the contribution of evaluation to budgeting is a key objective of the new evaluation policy, the new policy aims explicitly to build the role of evaluation in all relevant domains of public management. This includes an explicit role in program management improvement – in the words of the new policy, it aims to support "policy and program improvement" and "managing for results" as well as budgeting. In this sense, the new policy is similar to the Chilean strategy, which also aims to build an all-purpose evaluation function.

By contrast with the centralized Chilean approach, the new Canadian evaluation policy leaves evaluation as an essentially decentralized process. Both the management of specific evaluations and the choice of evaluation topics in any specific year remain essentially matters for the spending ministries themselves to decide. Most evaluations are carried out by internal evaluators within spending ministries, although outside evaluation experts support some evaluations (Treasury Board of Canada Secretariat 2011). The only qualification to the principle of decentralized control of evaluation is that the Treasury Board may

request that specific evaluation topics be included in ministry evaluation plans, although there have been few examples of such centrally mandated evaluations to date.

Implementation of the new evaluation policy took place progressively over four years, culminating in March 2013. It is therefore too early to assess its impact. Clearly the strong focus on reconnecting evaluation and budgeting, and the requirement that all evaluations focus strongly on effectiveness, are very welcomed. Nevertheless, the new policy raises a number of issues:

- Will the primarily decentralized approach to evaluation adequately serve the information requirements of central budget decision makers?
- When evaluation is left largely to the spending ministries, will the requirement that effectiveness issues be addressed overcome the past problem of primarily process-oriented evaluations, which have been of limited value for budgetary decisions?
- Will the center, in the form of the Treasury Board, play a sufficiently aggressive role in identifying programs it wishes to be evaluated to serve the budget preparation process?

The 2009 Canadian Evaluation Policy also raises the important issue of the desirability and practicality of a comprehensive evaluation strategy – that is, of a government-wide evaluation policy that mandates the evaluation of all government programs. In principle, such a comprehensive approach is highly attractive; in practice, it raises major resourcing issues. These were explicitly raised by the Auditor General in 2009.¹⁵

Both Chile and Canada stand out as examples of countries where the Ministry of Finance has taken the lead in implementing an evaluation strategy with the central aim of meeting the information needs of central budget decision makers. Both countries have made an effort to ensure that effectiveness remains a key focus of evaluations and to avoid the drift into overwhelmingly process-oriented evaluations. Nevertheless, both systems raise the issue of whether it is appropriate to combine budgetary evaluation system and government-wide evaluation policy into a single "all-purpose" evaluation strategy. The danger is that it may prove impossible to ride two horses and that budgetary requirements will end up subordinated to broader management improvement objectives.

¹⁵ Noting the "long-standing concern" about the "shortage of experienced program evaluators in the federal government," the Auditor General noted that "the departments we examined expressed concerns about their capacity to implement evaluation of all direct program spending, as required under the 2009 Policy on Evaluation. Even before these expanded requirements, they found it challenging to hire enough experienced evaluators to fully meet needs for effectiveness evaluation, and they had not been able to regularly address areas for improvement. In our view, identifying programs where effectiveness information can be put to the best use will be a key part of implementing the coverage requirements of this policy" (Auditor General of Canada 2009: 22, 33).

4. How to Better Connect Evaluation and Budgeting

Connecting budgeting and evaluation requires two things. First, evaluation needs to be made more useful for budgeting. Second, the budget process needs to focus more on expenditure prioritization and performance.

4.1 Making Evaluation More Useful for Budgeting

The starting point in considering how to make evaluation more useful for budgeting is to explicitly recognize that the information needs of central budget decision makers are significantly different from those of spending ministries. Broadly speaking, spending ministries seek to increase expenditure, while central budget decision makers—especially the ministry of finance—seek to control (and sometimes reduce) expenditures. Action to cut expenditures on existing programs—whether carried out to create additional "fiscal space" for important new spending or to reduce aggregate government expenditure—must generally come from the center and not from the spending ministries themselves.¹⁶ Conversely, new spending proposals—whether for entirely new programs, services, or projects or for additional funding for existing programs—will normally come from the spending ministries, and the role of the center is to decide which of the major new spending proposals should be accepted and which should be rejected.

This means that in respect to baseline expenditure, the potential contribution of evaluation is very different for central budget decision makers than it is for spending ministries. For the center, evaluation of baseline expenditure is important primarily as a means of helping to decide which programs to abolish or cut back.¹⁷ This makes outcome evaluations and program logic analysis the most relevant forms of evaluation, because these are the forms of evaluation which directly question the continued existence of programs by examining their effectiveness. Process evaluation has, for the center, a more limited and supportive role—primarily providing advice on whether ineffective programs are "fixable."

By contrast, for spending ministries the most useful role of evaluation of baseline expenditure is as a source of guidance on how to improve the design and management of programs. It is therefore process evaluations that are of greatest interest to spending ministries; outcome evaluations or program logic analysis tends to be seen as less useful. Moreover, the type of process evaluation that spending ministries find most useful is significantly different from that favored by the Ministry of Finance and other central budget decision makers. For spending ministries, process evaluations should mainly focus on managerial rather than budgetary issues and should assume rather than question the continued existence (even if in modified form) of the programs being evaluated. Although process evaluations carried out by spending ministries might play some budgetary role—in particular, helping to formulate requests for additional funding for existing programs to improve their effectiveness—any budgetary role is secondary.

¹⁶ Even if in principle spending ministries should continually question the justification for their ministry's programs, in practice this "challenge" function is not one they will readily assume.

¹⁷ Analogously, efficiency review is important to the center primarily as a means of deciding what budgetary cuts can be made based on improved efficiency.

In respect to proposals for new programs or services, there is less of a divergence of focus. At least in principle, both the center and the spending ministries have an interest in good ex ante evaluation (including program logic analysis) of proposals for new programs, services, or projects.

Against this background, it is useful to consider in more detail the type of evaluation most useful for central budget decision makers, focusing first on baseline expenditure and subsequently on ex ante evaluation.

4.2 What Type of Evaluation?

4.2.1 Outcome Evaluations and the Budget

As noted, for central budget decision makers, outcome evaluations and program logic analysis are of primary importance because they directly address program effectiveness, which should be a key criterion for targeting budget cuts. In principle, outcome evaluations should be the more important of these two because – unlike program logic analysis – they attempt to directly assess the effectiveness of programs. By contrast, program logic analysis "merely" asks whether the programs are theoretically capable of being effective.

It would, however, be wrong to say that the focus of budget-directed evaluations should be exclusively or even primarily on outcome evaluations. Outcome evaluations are insufficient for budgetary purposes because merely knowing whether a program is effective does not provide a sufficient basis for decisions on funding. Before cutting funding to an ineffective program, budgeters need to know whether the program can be *fixed*. That is, they need to know whether, without spending an unacceptable amount of additional money, changes in the design and/or management of the program could make the program effective. If it appears that an ineffective program might be saved in this manner, funding should not be immediately cut; the concerned ministry should instead be notified that it is expected to take steps required to reform the program and make it effective.

This means that where outcome evaluations indicate that a program is ineffective, it is necessary to conduct additional analysis to assess whether the program is "fixable" or essentially irredeemable and should therefore be considered as an option for budget cuts. What type of evaluation should be carried out to determine whether an ineffective program is "fixable"? In part, the answer lies in program logic analysis, which is directly relevant to the question of whether ineffective programs can be salvaged. If program analysis reveals that the program's mode of intervention is fundamentally flawed, it is probable that the program is incapable of being fixed and should instead be abolished.

This implies that outcome evaluations that are conducted for budgetary purposes should routinely be accompanied by program logic analysis. This highlights one of the three important roles that program logic analysis should play in supporting budgeting – the other two roles are an alternative to outcome evaluation and a core element of ex ante evaluation.

4.2.2 The Role of Process Evaluations

But what if the program logic analysis indicates that the program's strategy is broadly sound? Then the explanation for the program's lack of effectiveness may well lie in management or implementation problems. This does not, however, mean that the program should necessarily continue to be funded. Rather, the decision whether to continue or terminate funding should rest on an assessment of whether the spending ministry can actually fix the program's management or implementation problems at a reasonable cost and within a reasonable timeframe. If the answer is "yes," then funding should be continued at least on a provisional basis. But if the answer is "no," the program should be considered a prime target for funding cuts.

To make this assessment, a (specific type) of process evaluation may be useful. As suggested, however, this role should be viewed as limited and supportive. In particular, process evaluations to assess whether programs are fixable should be carried out on a highly selective basis. There is no value, from the central budget decision maker's perspective, in routinely combining outcome evaluations carried out for budgetary purposes with process evaluation focused on identifying opportunities for program improvement. It is only in cases where an outcome evaluation has found a program to be ineffective but a program logic analysis nevertheless suggests that the program strategy is sound that it may be useful to conduct a process evaluation to provide central budget decision makers with advice on whether the program can be saved through management/implementation changes. Conversely, it is wasteful to carry out this type of process evaluation for a program that an outcome evaluation has determined to be ineffective and that program logic analysis indicates to be fundamentally flawed. It follows that, for budgetary purposes, the first step should be to carry out outcome evaluation and/or program logic analysis, and then subsequently identify specific instances where it is useful to go further in carrying out a process evaluation before making the decision on continued funding.

Process evaluations designed to advise central budget decision makers on whether a program can be fixed through management/implementation changes will, moreover, tend to be different from the type of process evaluation normally carried out by spending ministries for program improvement purposes. The latter tends to assume the continued existence of the program and to identify potential improvement measures without weighing the cost of such measures against the improvements in program effectiveness they are likely to yield.

The need for budget decision makers to know whether an ineffective program is fixable should be reflected in the terms of reference given to the evaluators for this type of follow-up process evaluation. When a process evaluation is mandatory for budgetary purposes, the evaluators should not simply recommend options for improved program implementation, but should rather explicitly assess whether the spending ministry concerned is capable of substantially increasing the effectiveness of the program—at an acceptable cost.

Up to this point, the focus has been on the need to combine outcome evaluations with program logic analysis and, in selected cases, with a certain type of process evaluation. There are, however, important practical limits on the extent to which it is possible to carry out outcome evaluations to support budgeting.

4.2.3 Program Logic Analysis and Rapid Evaluation as a Substitute for Outcome Evaluation

It is very important for budget purposes that countries carry out more outcome evaluations. As noted, the lack of outcome evaluations is one of the key reasons for the historically disappointing contribution of evaluation to budgeting. Historic experience also highlights, however, that it is unrealistic to expect outcome evaluation to play a substantially expanded role until the set of *outcome indicators* is substantially improved. In certain advanced countries – the United Kingdom being an excellent example – there has been so much progress in developing outcome indicators over the past several decades that the absence of such indicators can no longer be regarded as an obstacle to the extension of outcome evaluation. Regrettably, this is not the case even in most other advanced countries. And in developing countries, outcome indicators tend to be relatively few and not always reliable.

It is therefore essential that countries continue to improve the quality, reliability, and coverage of outcome indicators as a prerequisite for expanding the role of outcome evaluations in supporting the budget. To the extent that good outcome indicators are not available in specific countries, though, the contribution of outcome evaluation to budgeting will necessarily be limited and the question will arise as to what role other forms of evaluation can play.

The availability of good outcome indicators is not the only constraint on the role outcome evaluations can play. Outcome evaluations are in general complex and costly and take significant time to prepare. They demand technical skills that are typically in short supply even in advanced countries. This means that everywhere –not only in developing countries – outcome evaluations need to be carefully rationed and focused on the programs where they can be of greatest assistance.

To the extent that it is impractical to carry out outcome evaluations of all programs that are considered questionable, program logic analysis plays an important role as a *substitute* for outcome evaluation. In other words, if a full outcome evaluation is impractical for reasons of cost or timeliness, a program logic analysis can be carried out to at least identify programs that should be considered for termination because of design flaws. Of course, such an analysis alone is incapable of providing the quality and reliability of information to budget decision makers that can the combination of a full outcome evaluation plus program logic analysis can provide. However, speed and cost considerations make it inevitable that the budget decision makers will often need to rely on program logic analysis alone.

Because it can be carried out quite quickly, program logic analysis is the core element of what is often called *rapid evaluation*. A rapid evaluation of a program can be defined as a combination of an analysis of the program logic with the analysis of any other data – such as outcome performance indicators – that may already be available and that may shed light on the program's effectiveness. Rapid evaluation relies on data that are already available and, unlike outcome evaluation, it does not include the collection of additional data to support analysis (World Bank 2004).

4.2.4 Evaluation of Program Relevance

From a budgetary point of view, the relevance of programs is as important as their effectiveness. As noted, "relevance" refers to the importance of the outcomes the program aims to achieve. A program may be effective in achieving its intended outcomes, but nevertheless find its funding terminated because the government considers it of low relevance. It is therefore an important part of good budgeting to continually reassess the relevance of existing programs.

In principle, such reassessment could be carried out as part of an evaluation program. Nevertheless, it can be argued that when analysis of the relevance of programs is necessary, it will normally be better for officials to do this – for example, Ministry of Finance staff or staff of some other relevant central agency – rather than professional evaluators. In other words, it should be officials who play the primary role in identifying programs that *appear* to be irrelevant and in advising the political leadership. The ultimate decisions about program relevance will then normally be made by the political leaders themselves.

Why should the analysis of program relevance primarily be the responsibility of officials rather than evaluators? First, judgments about program relevance are very closely linked with political philosophy and are a more subjective matter than program effectiveness. It is a key task of officials to understand the political philosophy of the government of the day and to provide analysis of program relevance that is attuned to the government's philosophy and values. Second, the analysis of program relevance is something in which well-trained government officials with strong policy analysis skills and long experience in government tend to excel.

4.3 The Selection of Evaluation Topics

When evaluation is intended to support the budget, the selection of evaluation topics as well as the type of evaluation carried out need to reflect budgetary priorities. Evaluation that is intended to assist central budget decisions should focus primarily on programs that are potential candidates for termination and where evaluation can help to decide whether funding should be withdrawn. This implies a high degree of targeting for budget-directed evaluation, rather than a broad-brush approach of seeking to evaluate all spending.

A number of points can be made about the selection of programs for evaluation to support budgeting:

First, the budgetary question will not usually be whether to abolish entire programs, but rather whether to terminate funding for specific components of programs (that is, subprograms or specific types of output within programs). Evaluations should therefore more generally be targeted at those components rather than at programs as a whole. Although this paper discusses the selection of programs for evaluation for budgetary purposes, this reference to programs rather than to subprograms or specific outputs is solely for reasons of brevity.

Second, there is no point evaluating programs that the government considers to be irrelevant or relatively unimportant. Such evaluation serves no useful purpose, as the government will, other things being equal, wish to terminate funding for such programs irrespective of their effectiveness.

Third, if a program is viewed as obviously ineffective, it is usually a waste of resources to conduct an evaluation to inform the decision as to whether to continue or terminate funding. Many programs of this type tend to exist across governments, usually because they were set up to buy votes rather than because they were seen as the best means of achieving specific outcomes. This will usually be obvious without an evaluation report.

Fourth – and qualifying the last point – it may nevertheless be useful to carry out a formal evaluation of an obviously ineffective program if the evaluation can help overcome political opposition to the abolition of the program. Demonstrating publically that a program is a waste of public money may help reduce the political cost of abolishing it.

Fifth, it serves no useful budgeting purpose to evaluate the effectiveness of programs that, no matter how ineffective they may currently be, provide services considered essential. For example, even if outcome evaluations of public primary schooling indicate that it is relatively ineffective in terms of the basic objectives of raising levels of literacy and numeracy, abolishing government primary schools is not an option. The relevant question is how to improve the quality of primary schooling – which is primarily a concern for the Education Ministry rather than the Ministry of Finance. Insofar as evaluations of effectiveness are intended to assist budget decision makers, they should therefore focus foremost on potentially expendable programs.

These points make it clear that the programs that should be selected for evaluation to assist expenditure prioritization will be different from the programs that should be targeted for program improvement purposes. For example, it makes good sense for program improvement purposes for spending ministries to carry out process evaluations of primary school education and other programs for which funding could never be withdrawn. For program improvement purposes, the range of programs for which it may be useful to carry out a process evaluation is broad, whereas evaluations designed to assist central budget decision makers need to be much more narrowly targeted.

4.4 Ex Ante Evaluation

All new spending proposals should be carefully analyzed to assess their cost effectiveness. Such analysis should be carried out in the first instance within the spending ministries that develop the proposals, and subsequently and in selected cases by the Ministry of Finance and any other relevant central agencies at the time that the new spending proposal is put forward for policy or budgetary approval. The analysis of new spending proposals is an essential function of the Ministry of Finance and one it can never delegate to other players. There should be well-defined routines requiring that such analysis be prepared and provided to the political leadership by the ministry for all substantial new spending proposals.

It is unimportant whether the rigorous analysis of new spending proposals by the Ministry of Finance and other central agencies is considered part of the evaluation program or whether it is labeled as policy analysis (with the evaluation label reserved for ex post evaluation). Unlike ex post evaluation, which will usually not be carried out by ministry of finance officials themselves, the analysis of new spending

proposals should be a key function of ministry of finance staff. Exactly how the responsibility for this analysis is assigned within the ministry of finance – whether it is the role of sectoral budget officers, or (as in Chile) of an evaluation unit – is a matter of choice. The discussion concerning the conduct of budget-linked evaluations therefore focuses on ex post evaluation.

4.5 The Conduct of Evaluations Linked to Budgets

The fact that the role of evaluation in supporting the budget process is very different from the role of evaluation in supporting spending ministry management has important implications for the organization and conduct of the program of evaluations carried out to support budget preparation. The Ministry of Finance cannot simply rely on evaluations carried out by spending ministries to provide the information that central budget decision makers need. The problem is not only that spending ministries have a natural interest in defending their budgets and may therefore give a positive "spin" to evaluations; more fundamentally, the selection of evaluation topics and their types of evaluations carried out by spending ministries cannot be expected to meet budgetary information needs.

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. This should be coordinated and planned together with a schedule of efficiency reviews that are similarly selected and managed to serve budgetary information requirements. This is referred to here as the *budgetary evaluation cycle*. For the most part, the Ministry of Finance will not itself carry out the evaluations or efficiency reviews, but will delegate these to independent evaluators and efficiency experts. However, the terms of reference and oversight of the evaluations and efficiency reviews should be firmly under the control of the Ministry of Finance. This does not, of course, rule out spending ministry participation, for example, via membership of a steering group set up to oversee evaluation of one of the ministry's programs.

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

- 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.
- Evaluations must be carried out quickly, so they do not lose pertinence – recall in this context the tight timelines that apply to evaluations in Chile. 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.
- 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 (for example, every three years).

4.6 Budget-Linked Evaluation and Government-wide Evaluation Policy

The need for a specific budgetary evaluation cycle – a program of evaluations and efficiency reviews specifically targeted and designed to serve the budget process – raises the question of what relationship such a cycle should have to any government-wide evaluation policy. As discussed, a government-wide evaluation policy refers to a centrally enforced policy that makes the conduct of evaluations mandatory for purposes including management improvement. As noted, Chile and Canada (together with Australia in the 1980s–early 1990s) are examples of countries where the Ministry of Finance has taken on the role of enforcing such a policy. In Canada (and, formerly, Australia), the evaluations carried out pursuant to the government-wide evaluation policy are carried out by the spending ministries themselves. In Chile, they are carried out under the control of the Ministry of Finance. In both countries, however, evaluations carried out as part of the government-wide policy were intended both assist budgeting and also support management improvement within the spending ministries.

Does the integration of budgetary evaluation and government-wide evaluation policy work? Together with the considerations raised in this section, experience casts doubts about this. Under a government-wide evaluation policy, the emphasis tends to be on broad coverage of evaluation topics, for example, in the form of a requirement that all programs are subject to evaluation over a certain timeframe. By contrast, selectivity is crucial to an effective budgetary evaluation cycle. This means that relying for budgetary purposes on whatever evaluations happen to be carried out in any specific years pursuant to the government-wide evaluation policy will tend to be unsatisfactory, because it will deny the Ministry of Finance the opportunity to target evaluations of the most questionable programs.

A further consideration is that exclusive reliance on evaluations carried out under the government-wide policy will mean that such evaluations need to be very broad in scope, routinely incorporating detailed process evaluation to meet the management improvement requirements of spending ministries *as well as* outcome evaluation and/or program logic analysis. The great danger – particularly if the control and management of evaluations is left to the spending ministries – is that in practice evaluations will focus more and more on the process evaluation component, with the consequence that the evaluations will tend to be of limited use for budget decision makers.

This suggests that, in countries where the Ministry of Finance assumes the role of implementing a government-wide policy aimed at ensuring full use of evaluation by spending ministries to improve management, the budgetary evaluation cycle should be kept distinct from the program of evaluations carried out pursuant to the government-wide evaluation policy.

4.7 Improving the Budget Process

Connecting – or reconnecting – budgeting and evaluation is, however, not merely a matter of modifying the nature of evaluation to make it more useful for budgeters. It is also a matter of changing budgeting itself to ensure that it makes full use of evaluation and other performance information.

4.7.1 A Focus on Priorities

This means transforming budget preparation so it focuses more on performance and expenditure prioritization. As noted, governments have all too often preferred to avoid expenditure prioritization because of political and bureaucratic sensitivities, or because of the perceived difficulty of the task. Clear program choices have been avoided partly through a willingness to allow aggregate government expenditure to continually increase. And when there has been a need for expenditure cuts, governments have often preferred to focus on cuts to input categories and to claim that efficiency gains alone can achieve the necessary reductions without the need to cut any services.

This approach is increasingly difficult to sustain under the fiscal circumstances that face many countries today. The unsustainable fiscal position that characterizes many countries as a result of decades of fiscal laxness, aggravated by the effects of the global financial crisis, makes substantial reductions in baseline expenditure unavoidable in the medium and longer terms. Recognition of this point has nothing to do with the debate about austerity versus growth. Even if one takes the view that excessively rapid fiscal consolidation is damaging in the short run, this in no way removes the necessity of such consolidation over time in countries that now have very high levels of debt and other financial liabilities and that face major adverse demographically related fiscal pressures such as rapidly growing pension expenditures.

In countries that need to cut public expenditure significantly, expenditure prioritization can no longer be avoided, and value for money in the delivery of public services has become critically important. Savings from efficiency are a crucial part of this, but it is no longer feasible to pretend that improved efficiency alone will be sufficient and that there is no need to identify effective and low-priority programs for termination or scaling back. It is for precisely this reason that performance budgeting is now a key focus of budget reform in many countries.

This points to the first change that needs to occur in budgeting if evaluation is to take its proper place as a budgeting tool: namely, that budget decision makers emphasize expenditure prioritization as the central element of the budget preparation process.

4.7.2 Spending Review as a Budgetary Routine

The second key change is the introduction of systematic *spending review* processes into the overall budget preparation process (Robinson 2013a). Spending review identifies options for significant cuts in baseline expenditure, whether from improved efficiency, the elimination of ineffective programs, or the elimination of programs that are low priority (that is, of low relevance). Spending review works best when it is a routine part of the budget preparation process carried out by officials (especially in the Ministry of Finance), as opposed to an ad hoc process carried out by outsiders (for example, prominent

businessmen). It should involve the creation of a spending review team within the bureaucracy with the explicit responsibility of assembling and putting forward options for cuts to baseline expenditure for consideration by the political leadership during the preparation of the annual budget. The spending review team does not carry out evaluations and efficiency reviews. Rather, it *uses* and *commissions* evaluations and efficiency reviews.

The spending review process is not merely a way to institutionalize better expenditure prioritization as part of the budget preparation process. It is also a means of creating a key "entry point" via which performance information including evaluation systematically enters into budget preparation. Without this, the budget process tends not to be geared to make use of performance information. The danger is then that, even if major efforts are made to develop performance indicators and evaluation, this information will not be properly used by budget decision makers.

Making spending review a continuing rather than ad hoc process is also crucial to building the role of evaluation because, as highlighted by the ad hoc Canadian "program review" of the mid-1990s and the French *révision générale des politiques publiques*, the short time frame of ad hoc reviews makes it hard to systematically commission evaluations of programs that the spending review team can use. With a continuing spending review process, the spending review team is able to commission evaluations and efficiency reviews that will be made use of the following year, or in two years' time.

4.7.3 Program Classification

The other key change required to facilitate the linkage of evaluation and budgeting is the adoption of a proper program classification of the budget. A good program classification identifies the outcomes and outputs upon which budgetary resources are spent, rather than simply the types of inputs or organizational units that absorb funds. Such a budget classification makes it possible for evaluation to explicitly target budgetary programs or, more usually, components of those programs (subprograms or specific types of service/outputs within the program). The explicit mapping of evaluation topics to budgetary programs makes it much easier to translate evaluation findings into changes in budget allocations.

Connecting evaluation and budgeting cannot, then, only involve supply-side changes to the nature of evaluation. It also requires demand-side changes to the nature of budgeting to ensure that budgeters recognize the importance of evaluation findings to the budget preparation process and are able to systematically make use of these findings.

5. Conclusions

Evaluation has a vital role to play in supporting better budgeting and is a core element of government-wide performance budgeting. The tendency of many to think of performance budgeting as being merely concerned with harnessing performance indicators to the budgeting process is a serious mistake. While good indicators are of fundamental importance, they also have major inherent limitations including the impossibility in most cases of distinguishing outcomes from the impacts of external factors by looking at outcome indicators alone. If budgeting is to become truly performance-informed, it is essential to build on good performance indicators through the systematic use of evaluation to analyze and interpret those indicators and thereby provide much better information about the effectiveness and efficiency of government spending.

Notwithstanding this, in the past evaluation has tended not to live up to its potential as a budgeting tool. This is not because nobody has made an effort to harness evaluation to budgeting. To the contrary, evaluation was seen as a key element of the information base supporting program budgeting when program budgeting systems were initially introduced in the United States and other countries in the 1960s and 1970s. However, neither at that time nor subsequently has the effort to connect evaluation and budgeting worked very well. Ministry of Finance officials, political leaders, and other key budget decision makers have tended to see evaluation reports as limited relevance the budget preparation.

Across the world, budget decision makers want this situation to change; they want evaluation to realize its potential as an instrument for improved budgeting. More than anything else, this is because of the tough fiscal circumstances which face many countries in the post-global financial crisis era. The need for medium-term fiscal restraint or fiscal consolidation is now widely recognized, even if the appropriateness of short-term austerity policies is actively debated. Under these circumstances, budgeting has become much more of a zero-sum game than in the past.

To be able to fund important new spending priorities increasingly requires cuts. "Spending reviews" to are increasingly an integral part of budget preparation, and it is widely recognized that evaluation has a potentially very important role to play in facilitating a more informed type of spending review. Specifically, evaluation can be of great assistance in helping spending review to identify programs which are irredeemably ineffective or which do not provide results sufficient to justify the resources they absorb.

To bring evaluation and budgeting together requires that they both change. Insofar as it is intended to support better budgeting, evaluation must be tailored to that purpose through the appropriate selection of evaluation topics (that is, focusing on areas of potential budgetary savings). The type of evaluation carried out also needs to reflect budgetary concerns, particularly by helping identify savings options rather than focusing on management and policy improvements, which has tended to dominate evaluation in the past. This means more focus on outcome evaluation and program logic analysis.

The need to ensure that evaluations are tailored to the needs of budget decision makers also raises the question as to whether government-wide evaluation systems should be expected to deliver the types of

evaluations budget decision makers need, or whether it may be appropriate to operate a budgetary evaluation cycle separate from any government-wide evaluation system.

Glossary

Activity: A work process by which inputs are used to produce outputs.

Aggregate fiscal policy: A government's overarching objectives for the budget deficit, debt, and other relevant fiscal aggregates.

Budget classification: Categories of expenditure used in the budget, particularly for the approval of expenditure.

Budget preparation: Stage of the budget process in which the government decides how much funding it will provide to which agencies and for which purposes, and prepares the budget law that will be presented to parliament for its approval.

Cost-effectiveness: The achievement of intended outcomes at the lowest possible cost.

Effectiveness: The degree of success of an output in delivering its intended outcomes.

Efficiency: Production of an output at minimum cost while holding quality constant, given prevailing input prices.

Evaluation: Analytic assessments typically addressing the cost-effectiveness or appropriateness of policies, organizations, or programs.

External factors: Factors outside the control of government that influence the outcomes achieved by public programs. External factors may be either client/case characteristics or aspects of the context in which the program is delivered. Sometimes also referred to as contextual factors.

Formula funding: When used as a performance budgeting tool, a system in which funding provided by a government to a public sector agency is an explicit function of measures of expected and/or actual results—that is, of measures of outputs and/or outcomes.

Impacts: Term used to refer to longer-term or higher-level outcomes. In this paper, incorporated within the concept of *outcomes*.

Incrementalism: Budgeting is characterized by “inattentiveness to the (budgetary) base”—in other words, that budgetary decision makers take the budgetary base more or less for granted as the starting point in budget formulation, and focus their attention primarily on the size of the increment (or, occasionally, decrement) in agency or program budgets, mainly by a process of adjusting budgets for cost changes.

Inputs: Resources used in carrying out activities to produce outputs (for example, labor, equipment, buildings).

Line-item budgeting: Budgeting in which agencies are provided with budget appropriations specified in terms of input categories (that is, by economic classification).

Outcome: Changes brought about by public interventions on individuals, social structures, or the physical environment.

Output: A good or service produced as a result of activities, and provided by the ministry to an *external* client.

Performance indicator: See *performance measure*.

Performance measure: Ratings or other quantitative measures that provide information on the effectiveness and efficiency of public programs or processes.

Processes: The processes by which inputs are transformed into outputs. Same as *activities*.

Program budgeting: The systematic use of performance information to inform decisions about budgetary priorities between competing programs, based on the program classification of expenditure (see *programs*).

Programs: Categories of expenditure primarily based on groups of outputs (or support services) with a common objective, which is usually an outcome.

Purchaser-provider system: Funding systems under which agencies are paid for the results (usually outputs) they deliver.

Quality: The extent to which the characteristics of an output—in the case of a service output, the activities delivered and their timeliness— are such that they increase its potential capacity to achieve its intended outcome. Not to be confused with the outcome itself.

Spending review: The systematic scrutiny of existing expenditure to identify, in particular, options for cuts. It involves both the review of specific services provided by government and efficiency reviews, which focus on reducing the cost of delivering services.

Strategic phase (of the budget preparation process): Early stage in a budget preparation process where the political leadership sets the priorities that will guide the process.

Transfers: Shifting of funds between appropriation categories.

References

- Arenas, A., and H. Berner (2010) *Presupuesto por Resultados y la Consolidación del Sistema de Evaluación y Control de Gestión del Gobierno Central* (Santiago de Chile: Ministerio de Hacienda).
- Aucoin, P. (2005) "Decision-Making in Government: The Role of Program Evaluation Discussion Paper." <http://www.tbs-sct.gc.ca/cee/tools-outils/aucoin-eng.asp#fn06>.
- Auditor General of Canada (1993) *1993 Report of the Auditor General of Canada*. http://www.oag-bvg.gc.ca/internet/English/parl_oag_199311_e_1157.html.
- Auditor General of Canada (2009) *2009 Report of the Auditor General of Canada* (Ottawa: Office of the Auditor General of Canada).
- Bourgon, J. (2009) *Program Review: The Government of Canada's Experience Eliminating the Deficit, 1994-99: A Canadian Case Study* (London: Institute for Government).
- Chevauchez, B. (2007) "Public Management Reform in France." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Department of Finance and Australian Public Service Board (1987) *Evaluating Government Programs: Financial Management Improvement Program* (Canberra: Australian Government Publishing Service).
- Floden, R.E., and S.S. Weiner (1982) "Rationality to Ritual: The Multiple Roles of Evaluation in Government Processes." In F.J. Lynden and E.G Miller *Public Budgeting*, 4th ed. (Englewood Cliffs, NJ: Prentice-Hall).
- GAO (Government Accountability Office) (2011) *Program Evaluation: Experienced Agencies Follow a Similar Model for Prioritizing Research*, GAO-11-176. (Washington, DC: GAO).
- Good, D. (2008) *The Politics of Public Money* (Toronto: University of Toronto Press).
- Gouvernement de France (1998) *Circulaire du 28 décembre 1998 relative à l'évaluation des politiques publiques*. <http://legifrance.gouv.fr>.
- Guzman, M. (2007) "The Chilean Experience." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Hawkesworth, I., O. Huerta Melchor and M. Robinson (2012) "Selected Budgeting Issues in Chile." *OECD Journal on Budgeting* (2):147-185.
- Huet, P. (1971) "Présentation de la Rationalisation des Choix Budgétaires en France." In P. Huet et al. (eds.), *Rationalisation des Choix Budgétaires: Vers Une Nouvelle Raison d'État* (Paris: La Documentation Française).
- Lahey, R. (2010) *The Canadian M&E System: Lessons Learned from 30 Years of Development* (Washington, DC: World Bank).

- Lannaud, B. (2007) "Performance in the New French System." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Marvin, K., and A. M. Rouse (1970) "The Status of PPB in Federal Agencies: A Comparative Perspective." In R. H. Haveman and J. Margolis (eds.), *Public Expenditures and Policy Analysis*. (Chicago: Markham Publishing Company).
- Mackay, K. (1998) *The Development of Australia's Evaluation System*. ECD Working Paper No.4. (Washington, DC: World Bank).
- Mackay, K. (2004) *Two Generations of Performance Evaluation and Management System in Australia*, ECD Working Paper No. 11. (Washington, DC: World Bank).
- OECD (1998) *Best Practice Guidelines for Evaluation*. PUMA Policy Brief No. 5. (Paris: OECD).
- OECD (2000) *Glossary of Key Terms in Evaluation and Results Based Management* (Paris: OECD).
- OMB (2002) *Instructions for the Program Assessment Ratings Tool*. (Washington, DC: OMB)
- Orszag, P. (2009) *Memorandum for the Heads of Executive Departments and Agencies: Increased Emphasis on Program Evaluations*, M-10-01. (Washington, DC: Office of Management and Budget).
- RGPP (Révision Générale des Politiques Publiques) (2011) *5ème Conseil de Modernisation des Politiques Publique, mars 2011: Sommaire*. <http://www.modernisation.gouv.fr/>.
- Robinson, M. (ed.) (2007a) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Robinson, M. (2007b) "Performance Budgeting Models and Mechanisms." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Robinson, M. (2007c) "Purchaser-Provider Systems." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).
- Robinson, M. (2011) *Performance-Based Budgeting* (Washington, DC: CLEAR/World Bank).
- Robinson, M. (2013a) *Spending Review*, GOV/PGC/SBO(2013)6 (Paris: OECD).
- Robinson, M. (2013b) *Program Classification for Performance-Based Budgeting: How to Structure Budgets to Enable the Use of Evidence*. (Washington, DC: World Bank).
- Robinson, M. (2013c) "Performance Budgeting." In R. Allen et al. (eds.) *International Handbook of Public Financial Management* (Houndsmills: Palgrave Macmillan).
- Smith, P. (2007) "Formula Funding and Performance Budgeting." In M. Robinson (ed.) *Performance Budgeting: Linking Funding to Results* (Basingstoke and New York: Palgrave Macmillan/IMF).

Steiner, G.A. (1967) "Problems in Implementing Program Budgeting." In D. Novick (ed.) *Program Budgeting: Program Analysis and the Federal Budget* (Cambridge, MA: Harvard University).

Toulemonde, J., and L. Rochaix (1994) "Rational Decision-Making through Project Appraisal: A Presentation of French Attempts." *International Review of Administrative Sciences* (60): 37–53.

Treasury Board of Canada Secretariat (2004) *Study of the Evaluation Function In the federal Government*. Prepared by the Centre of Excellence for Evaluation (Ottawa: Treasury Board of Canada Secretariat).

Treasury Board of Canada Secretariat (2009) *Policy on Evaluation*. <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024§ion=text>.

Treasury Board of Canada Secretariat (2011) *Expenditure Management: The Canadian Perspective*, Presentation to World Bank Independent Evaluation Group's Evaluation Week, October 25.

Weinstock, M. (2003) "Under the Microscope." *Government Executive* (January): 37–40.

Wholey, J.S. (1978) *Zero-Base Budgeting and Program Evaluation* (Lexington, MA: Lexington Books).

World Bank (2004) *Monitoring and Evaluation: Some Tools, Methods and Approaches* (Washington, DC: World Bank).

Wye, C. (1989) "Role of the Evaluation Function in Federal Government." *American Journal of Evaluation* 10(2): 24-25.