The Role of Evaluation in Spending Review

Marc Robinson
PFM Results Consulting

Abstract: The role of spending review is to identify savings options that enable governments either to find fiscal space for priority new spending or to cut aggregate spending. Spending review has been extensively used by governments around the world in the wake of the global financial crisis in 2008, and many governments are now seeking to institutionalize spending review as a permanent part of the budget preparation process. The effectiveness of spending review is critically dependent upon the quality of its information base—that is, of the expenditure analysis and performance indicators that can assist in the search for savings options. Evaluation is an essential part of this information base. However, ensuring that the potential of evaluation to inform spending review is realized will require considerable reflection on the design, selection, and conduct of evaluations.

Keywords: baseline expenditure, budget preparation process, fiscal space, information base, savings options

Résume: L'objectif des exercices de révision budgétaire est d'identifier des moyens d'économiser afin que les gouvernements puissent soit trouver une disponibilité fiscale pour de nouvelles dépenses, soit réduire les dépenses globales. À cause de l'utilisation mondiale accrue d'exercices de révision budgétaire qui a suivi la crise financière de 2008, de nombreux gouvernements cherchent à institutionnaliser ces exercices de façon permanente dans le processus de préparation des budgets. L'efficacité des révisions budgétaires repose de façon critique sur la qualité de l'information – c'est-à-dire l'analyse des dépenses et les indicateurs de performance qui peuvent aider à trouver des sources potentielles d'économie. L'évaluation est une source importante d'information. Toutefois, pour que les évaluations informent les exercices de révision budgétaire nous devons d'abord réfléchir sur la façon dont les évaluations sont conçues, orientées et réalisées.

Mots clés : dépenses, processus de préparation du budget, espace budgétaire, base de données, options d'épargne

WHAT IS SPENDING REVIEW?

Spending review is a process for systematically scrutinizing baseline expenditure to identify and implement savings measures (Robinson, 2013). Savings measures are

Corresponding author: Marc Robinson, Bosquets de Paudille 16, Chardonne, Switzerland; marc@pfmresults.com

specific actions, such as terminating a program or streamlining a business process, that government can take to reduce baseline expenditure. *Baseline expenditure* means expenditure on existing programs and projects, at levels required by prevailing policies or laws. It does not include any expenditure on new initiatives (or on the deliberate expansion of existing programs). The appraisal of new spending proposals is therefore not spending review. Spending review focuses on baseline expenditure, because in most countries the budget preparation process focuses almost exclusively on the consideration of new spending proposals and fails to seriously look at the justification for the continued funding of baseline expenditure. Spending review is therefore an instrument for attacking budgetary *incrementalism*—that is, the tendency for baseline funding to be unthinkingly renewed in each budget.

A spending review may be tasked to identify one or both of two types of savings measures: strategic savings and efficiency savings. Strategic savings are expenditure reductions achieved by cutting back services (outputs) delivered to the community—in other words, by scaling back or terminating programs. Strategic savings might, for example, target programs that are ineffective or low priority. Efficiency savings, by contrast, are expenditure reductions that are achieved by changing the way in which outputs are produced so as to deliver the same quantity and quality at lower cost. Today, the majority of countries with spending review systems use it to make both strategic and efficiency savings.

To identify savings options, spending reviews typically apply several review criteria. The main review criteria are relevance, duplication, effectiveness, and efficiency. In reviewing some programs, equity and market failure are pertinent additional criteria.

THE OBJECTIVES OF SPENDING REVIEW

One of the main reason governments use spending review is to reduce, or restrain the growth of, aggregate expenditure. In the wake of the global financial

Table 1. Review Criteria Employed during Spending Review to Identify Savings Options

Relevance	Is the program trying to achieve an outcome for society that is important?
Duplication	Is there duplication with another program or agency in govern- ment, or with the private sector?
Effectiveness	Is the program achieving its intended outcome, and, if not, can it be fixed so that it does so?
Efficiency	Could the service or process be produced at lower cost, without compromising quality?
Equity	Is the distribution of the service or transfer payment between different categories of recipient equitable?
Market failure	Is government doing something that could be left to the private sector or community organizations to do?

crisis, certain governments considered that they needed to implement large fiscal consolidations, with large cuts in aggregate expenditure (Robinson, 2015). These governments found spending review to be an invaluable tool to achieve these cuts. They used the savings realized through spending review entirely or largely to reduce the budget deficit rather than to finance new spending priorities. The United Kingdom and Ireland (Boyle, 2011) are particularly good examples of governments that made use of spending review to achieve major fiscal consolidations.

Governments have found that, faced with the need to make large reductions in expenditure, it is far better to do this in a targeted way through spending review than to rely upon crude across-the-board cuts. Across-the-board cuts not only fail to discriminate between high-priority and low-priority spending, but they are often implemented by spending ministries in a way which adversely affects effectiveness (e.g., always cutting capital expenditure and the most flexible elements of operating expenditure, such as supplies and training).

Spending review is also used by governments to increase the *fiscal space* available for high-priority new spending initiatives. In most countries, it is quite difficult to find funds for new spending while keeping aggregate government expenditure and the budget balance at reasonable levels. The weight of baseline expenditure requirements inherited from the past tends to continually increase, while revenue growth is in most countries slower these days than it was in the boom years before the global financial crisis. Under these circumstances, making savings by cutting back ineffective or low-priority programs, or by improving efficiency, can substantially increase governments' freedom of movement in responding to new spending pressures. By facilitating the reallocation of money from baseline expenditure to new priorities, spending review improves the overall prioritization of government expenditure. Resources are allocated more to where they are needed and less to where they happen to have been allocated in past years. Overall value for money increases.

Spending review should not, thus, be regarded only as an instrument for fiscal consolidation—that is, as an axe wielded by austerity-obsessed governments. In the longer run, spending review is better viewed primarily as a means of increasing fiscal space and improving expenditure prioritization. This is why many governments are seeking to institutionalize spending review as a routine ongoing process rather than as an exercise to be carried out only in crisis conditions.

SPENDING REVIEW DESIGN FEATURES

Prior to the global financial crisis, only a handful of countries (e.g., Denmark and the Netherlands [van Nispen, 2016]) had operated spending review systems on a continuing basis. Some countries (including Canada in the 1990s [Good, 2008; Bourgon, 2009]) had in the past used spending review as a one-off tool, but had abandoned it.

Since the global financial crisis, spending review has become a much more important tool of government budgeting. In the tighter fiscal context facing governments since that time, there has been a rapid increase in the number of countries making systematic use of spending review, reflecting the prevailing tighter budgetary circumstances facing governments. One distinctive feature of spending review today is that it is generally designed as an ongoing, rather than one-off, process. This means that spending review is conceived of as a process to be undertaken on a regular basis into the indefinite future. "Regular" may mean annual, or it may mean at multi-year intervals (e.g., every 3 or 4 years).

Another closely related feature of contemporary spending review is that, in most countries, it is intended to be an integral part of the budget preparation process. This means that spending review processes are in general deliberately designed as mechanisms to feed savings options to the government for consideration and decision in an upcoming budget—that is, as part of the overall process of deciding how much budget funding each spending ministry will receive. Integration of spending review into the budget preparation process has implications for the frequency of spending review. It means that in countries with fixed multi-year budgeting systems—such as the UK, which essentially prepares every 3 years a budget for the 3 coming years—it makes no sense to conduct spending review as an annual process. On the other hand, countries with annual budgeting or with rolling multi-year budgets² may choose either to carry out some spending review every year or to undertake it on a multi-year schedule.

The contemporary conception of spending review as an ongoing process integrated into budget preparation contrasts with the approach taken in the past by several governments (e.g., during the 1980s) that set up ad hoc spending reviews (e.g., carried out by independent commissions) which were quite separate from the budget process. The integration of spending review into the budget preparation process has the important consequence that the process of developing, presenting, and deciding to adopt savings options must fit into the tight time constraints of the preparation of the budget. As discussed below, this has implications for the role of evaluation in spending review. There has been a tendency over recent decades for the budget calendar—the sequence of steps and dates that comprise the budget preparation process—to become longer. Today, the time from the commencement of the budget preparation process through to the finalization of the draft budget for presentation to the parliament may take up to 9 months. However, most of the decisions about savings options to be incorporated in the budget will need to have been taken several months before the end point of this process.

There are important differences between nations in the design and scope of the spending review process (Robinson, 2013, 2015), not all of which are relevant to the issue of the role of evaluation. One difference that is, however, of some consequence for the role of evaluation is that between selective and comprehensive spending reviews. A *selective* spending review is a spending review that is limited to several predefined review topics. For example, the government might decide that in the coming year the spending review process will look for savings options from review topics such as (a) social assistance benefits to families, (b) information and communications technology acquisition and management, and (c) waste management services. Denmark and the Netherlands (van Nispen & de Jong,

2017)³ are examples of countries that use the selective spending review approach, choosing anything up to 20 review topics in each round of spending review.

A comprehensive spending review, on the other hand, is a spending review process that is not limited to any predefined list of review topics. In a comprehensive spending review, the ministry of finance and spending ministries are asked to undertake an unconstrained search for the best savings options. This does not mean that in a comprehensive spending review all programs and business processes are reviewed—that would be impossible. Countries that have undertaken comprehensive spending reviews include the UK and Ireland.

EVALUATION AND THE INFORMATION BASE OF SPENDING REVIEW

The information base of spending review refers to information, such as performance indicators and expenditure analysis, that is available to those carrying out the spending review process and that is potentially of assistance to them in applying the review criteria to identify savings options. A range of different types of expenditure analysis may form part of the spending review information base. For example, economic analysis is particularly useful when applying the market failure criterion. Evaluation is, however, of particular importance. It is perhaps not useful in this article to concern ourselves too much with the distinction between evaluation and other types of expenditure analysis (other than to say that it is perhaps not terribly useful to use the term evaluation to cover every form of expenditure analysis). It is sufficient, rather, to say that, in the senses that the term evaluation is most commonly used, it is potentially of greatest relevance in applying the effectiveness criterion in the search for strategic savings options. If, however, the term evaluation is also used to refer to efficiency analysis, then evaluation is clearly also an important part of the information base in the search for efficiency savings.

Evaluations do not per se constitute spending reviews. 4 Spending reviews happen only when working groups undertake the task of identifying explicit savings options. This is something that evaluations do not necessarily (nor even typically) do.

Particularly in the immediate aftermath of the global financial crisis, most spending reviews carried out by OECD governments were of the "quick and dirty" variety, with little use of evaluation or other systematic forms of expenditure analysis. Many ministries of finance, as leaders of the spending review process, felt that this was unsatisfactory and decided that strengthening the information base of spending review was an important priority if spending review was to be institutionalized as an ongoing process linked to budget preparation. A key aspect of this, in their view, is the strengthening of evaluation and the more systematic linking of evaluation to spending review.

If evaluation is potentially of considerable value in strengthening spending review, a number of key questions must be addressed before that potential value can be realized in practice. One is the question of the procedural link, if any, between the spending review process and the conduct of evaluations. Should evaluations be integrated into spending review as part of a single process, so that, for example, a spending review of a specific program should (where appropriate) include evaluation of the program? Or should evaluation instead be a separate process, carried out before, and independently of, spending review? Should spending review simply opportunistically make use of pre-existing evaluations? Or should evaluations sometimes be specifically commissioned to serve as part of the information base for subsequent spending review?

EVALUATION AND THE PROGRAM EFFECTIVENESS CRITERION

Applying the effectiveness criterion to the search for savings options means identifying programs that are either completely ineffective (i.e., they fail totally to achieve their intended outcomes) or which are not cost-effective (i.e., they achieve their intended outcomes to some extent, but their cost is excessive relative to the outcomes achieved). Evaluation is potentially of great relevance in establishing the effectiveness of programs. As is well known, effectiveness cannot in general be assessed by reference to outcome indicators alone, even when the best possible outcome indicators have been developed. The majority of outcome indicators fail to (fully) distinguish between the true impact of government programs and the impact of external factors.⁵

In principle, the type of evaluation that can be of greatest assistance in applying the efficiency criterion in spending review is impact evaluation. Good impact analysis can be of particular assistance in helping to distinguish the effectiveness of the program from the impact of external factors. It might, for example, apply multivariable regression analysis to time series data for this purpose. Nevertheless, impact evaluation alone cannot usually be expected to tell decision makers which programs should be cut or wound back. While it is sometimes the case that an impact evaluation will indicate that a program is completely ineffective—that is, that it appears to achieve no outcomes at all—this is usually not the case. In most cases, impact evaluations find that programs achieve their intended outcomes to some extent. The key question then becomes that of cost-effectiveness. The outcomes achieved need then to be assessed against the cost of the program. Comparing benefits with costs is not something that evaluation itself usually does. The main formal technique for doing this is, of course, benefit-cost analysis. However, in practice, judgements of whether outcomes are sufficient to justify cost are usually made informally rather than by deploying benefit-cost analysis-both because the conclusions reached by benefit-cost analysis are often not very robust and because benefit-cost analysis is often an expensive and time-consuming process.

Notwithstanding the potentially considerable value of impact evaluations as part of the information base for spending review, there is a major practical constraint on their use. This is the time that it normally takes to conduct a proper impact evaluation—which is typically at least a couple of years. This makes it

impossible to initiate an impact evaluation and use its results to help identify savings options all within the time frame of the budget preparation process. This means that impact evaluations must necessarily be carried out prior the budget preparation cycle in which they are used as an input to a spending review. This can happen in one of two ways. The first is that a program might be identified as the subject of a spending review in a future budget preparation cycle (e.g., in 2 or 3 years' time), and an evaluation initiated now with a deadline tailored appropriately. When this approach is taken, it is relatively easy to design the evaluation to meet the needs of spending review.

The other approach is to have an across-the-board evaluation policy, which requires that all programs (or all major programs) be subject to periodic evaluations. Such across-the-board evaluation policies have, of course, existed at various stages in several countries (including Canada, Australia, and Chile). Evaluations carried out pursuant to an across-the-board evaluation policy are usually intended to have multiple purposes and are unlikely to be designed specifically to serve the needs of spending review. This raises the question—discussed further below—of the extent to which evaluations are designed to provide useful information for spending review.

The difficulty of carrying out impact evaluations quickly is one of the factors that have led to recent emphasis among evaluators on the role of rapid evaluations. Rapid evaluations may use several different methodologies, but program logic analysis (sometimes also known as theory-based evaluation [World Bank, 2004, p. 10]) can be considered to be one of its core methodologies. 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 (a.k.a. intervention 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 and existing lines of evidence. Program logic analysis is, in a sense, one of the most standard tools of government analysts, and it is open to debate whether it deserves to be labelled evaluation. What is perhaps distinctive about it as an instrument of rapid evaluation is probably not so much the methodology itself as the degree of rigor and formality with which it is applied.

Unlike impact evaluations, rapid evaluations—including rigorous program logic analysis—can, where appropriate, be carried out as an integral part of the spending review process. This does not, however, mean that they should be a mandatory part of all strategic spending review. If, for example, a program is regarded as questionable primarily on the grounds of relevance, it may be quite pointless to carry out any type of evaluation of its effectiveness.

EVALUATION AND EFFICIENCY SAVINGS

Identifying options for efficiency savings also requires both better performance indicators and the right type of analysis. By definition, efficiency improves when cost per unit of output (service) falls, holding the quality of output constant. Output unit cost indicators are therefore often useful in identifying potential efficiency savings. Time series analysis and benchmarking (i.e., cross-sectional comparisons) of unit costs can be particularly revealing. There are, however, several potential problems with relying exclusively on unit cost indicators. One is that factors other than efficiency can—and often do—impact on output unit costs. For example, benchmarking may show that the unit costs of providing primary health services are significantly higher in rural areas than in cities, but this may simply be because geography and lower population density make it inherently costlier to deliver services outside cities. In this case, higher unit costs do not imply that there is an opportunity for efficiency savings.

For this and other reasons, efficiency analysis—which we will take here to be another form of evaluation, even if it is often carried out by analysts who do not label themselves as evaluators—often plays an essential role in identifying potential efficiency savings. Methodologically, efficiency analysis can take several forms. Business process analysis—which aims to identify ways of streamlining the service delivery process—is one. Another approach is cost factor analysis, which involves taking unit cost indicators and adjusting them for the impact of cost disability factors (e.g., the higher cost of primary health service delivery in rural areas with lower concentrations of population).

It is more difficult to generalize about the time necessary to undertake efficiency analysis than it is to generalize about time necessary to undertake, say, impact evaluations. In many cases, it is perfectly feasible to carry out efficiency reviews of some specific business processes or programs within the time frame of the budget preparation process. Even wide-ranging major efficiency reviews—such as the 2003–2004 Gershon review, carried out in the UK (Gershon, 2004)—have sometimes been conducted within time frames which, even if longer than that of a single budget preparation cycle, have been quite short. On the other hand, in the case of some business processes, it is wise to spend more time on the review process.

This means that efficiency analysis may in some cases be carried out as an integral part of the spending review process, but that in some cases it should be carried out over a longer time frame in preparation for a future spending review cycle.

THE RELEVANCE OF EVALUATION FOR SPENDING REVIEW

It is clear that, in principle, evaluation should be a key part of the information base of spending review. In particular, it should assist reviewers greatly in applying the effectiveness and efficiency criteria to the search for savings options. In practice, however, the contribution of evaluation to spending review has not, at least up to recently, realized this potential. Interviews conducted with senior budget officials

in OECD countries in early 2013 (Robinson, 2013) found that, without exception, they considered that evaluations had made little contribution to the spending review process in their countries. Strikingly, this view was shared both by officials from countries where there were strong evaluation systems (e.g., Canada) and by those from countries where evaluation practice was weaker (e.g., France).

The reasons for this failure to connect evaluation and spending review (and, more generally, evaluation and budgeting) have been explored in detail elsewhere (Robinson, 2014). In summary, key factors appear to include

- the tendency of much evaluation to be focused on management/policy improvement;
- the desire of evaluators to be seen as helpful to managers, and consequently to avoid framing evaluation reports in a way that may be seen as too harsh or critical;
- the practice in some countries of allowing evaluations to be primarily controlled by the ministries that are being evaluated, at the risk of their independence;
- the tendency, as a result of some or all of the above factors, to focus primarily on making "positive" management/policy improvement recommendations;
- the difficulty of making definitive judgements—for example, about program effectiveness—in the context of imperfect information/high information costs, and the unwillingness in some cases of evaluators to use their "best judgement"; and
- the tendency for the choice of evaluation topics not to give much weight to the potential for budgetary savings.

Despite concerns about limited usefulness of evaluations to date, ministries of finance in many OECD countries were at the time of the 2013 interviews keen to strengthen the role of evaluation as a tool for budget decision-making. Most felt that, in the absence of a strong information base, spending review had been far too "seat of the pants" and needed to be made more rigorous. The extent to which action has been taken to strengthen the information base of spending review since that time is an important research topic.

An important associated issue is that of administrative responsibility for evaluations that are intended to support budget decision-making, including spending review. It is probably not realistic to expect that evaluations can be all things to all clients. It may therefore be entirely appropriate to maintain evaluation systems that are primarily focused on management/policy improvement and that are as far as possible decentralized in the management, while at the same time endowing ministries of finance with an evaluation service capable of carrying out evaluations the explicit objective of which is to assess value-for-money and to provide advice for those who need to decide whether to cut (or increase) the budgets of specific programs or agencies.

That evaluation has the potential to make a much bigger contribution to spending review does not, it should be stressed, mean that the identification of saving options in the spending review depends exclusively on evaluation. For example, the identification of programs that do not pass the "relevance" test does not require evaluation, and relevance is often a question of political philosophy. Decisions on savings measures are ultimately political, and not merely technical. However, better technical analysis to identify savings measures—including through evaluation—can help political leaders considerably in making more and better-targeted savings.

CONCLUDING OBSERVATIONS

It is important for the future of spending review to develop a better information base to assist reviewers to identify appropriate savings options. Evaluation, however defined, is a crucial part of that information base. However, experience shows that merely undertaking more evaluations will not ensure their use in spending reviews. For evaluation to deliver its potential benefits for spending review, it is important that the selection of review topics, and the design of the evaluations themselves, reflect the needs of budget decision makers. If it is unrealistic to think that all evaluations should serve in equal measure the needs of all potential clients, this may mean that there is a need to create an evaluation stream that is specifically dedicated to meeting the needs of budget decision makers.

The relevance of all of this depends, ultimately, on the success of government and ministries of finance in institutionalizing spending review as an ongoing part of the budget preparation process. The case for doing so is impeccable, and there is a strong feeling in many countries that this should happen. At the same time, however, enthusiasm for spending review has waxed and waned in the past, and the jury remains out on whether this time will be truly different. If it is, then it will be essential to ensure that evaluation makes the maximum possible contribution to the spending review process.

NOTES

- 1 Efficiency savings pertain to what economists refer to as *productive* efficiency, and not to other efficiency concepts such as *allocative* efficiency.
- 2 This means that the budget is formulated in, say, a 3-year time horizon, but is updated each year.
- 3 Although the government described the spending review carried out in 2010 in the Netherlands as a "comprehensive" review, this merely meant that a large number of predefined review topics were considered. It was therefore a selective, rather than comprehensive, review, as these terms are used in this paper.
- 4 This view of the relationship between evaluation and spending review differs from that put forward by, for example, Vandierendonck (2014).
- 5 Equally well known is the fact that, for certain types of outcomes, no outcome measurement metric exists (e.g., to measure the degree of national security).
- 6 Another important problem is that measuring output unit cost is also often technically challenging. Hence, relying on performance indicators alone to assess the scope for

efficiency savings has its limitations. In practice, it may be more effective to do a costing exercise selectively (i.e., only for services being reviewed) rather than develop acrossthe-board efficiency measures for all services.

REFERENCES

- Bourgon, J. (2009). Program review: The government of Canada's experience eliminating the deficit, 1994-99: A Canadian case study. London, UK: Institute for Government.
- Boyle, R. (2011). Better use of public money: The contribution of spending reviews and performance budgeting (Research Paper No. 5). Dublin, Ireland: Institute of Public Administration.
- Gershon, P. (2004). Releasing resources to the front line: Independent review of public sector efficiency. London, UK: HMSO.
- Good, D. A. (2008). The politics of public money: Spenders, guardians, priority setters, and financial watchdogs inside the Canadian government. Toronto: University of Toronto Press.
- Robinson, M. (2013). Spending reviews. OECD Journal on Budgeting, 13(2), 81-122.
- Robinson, M. (2014). Connecting evaluation and budgeting (ECD Working Paper Series No. 30). Washington, DC: World Bank.
- Robinson, M. (2015). Revisões de despesas na OECD. In R. Boueri, F. Rocha, & F. Rodopoulos (Eds.), Avaliação da qualidade do gasto público e mensuração da eficiência (pp. 107-160). Brasilia, Brazil: Tesouro Nacional.
- van Nispen, F. (2016). Policy analysis in times of austerity: A cross-national comparison of spending reviews. Journal of Comparative Policy Analysis, 18(5), 479-501. https:// doi.org/10.1080/13876988.2015.1005929
- Van Nispen, F., & de Jong, M. (2017). Evidence-based budgetary policy: Speaking truth to power? (EUI Working Papers, SPS 2017/1). Retrieved from the European University Institute, Department of Political and Social Sciences website: http://cadmus.eui.eu/ bitstream/handle/1814/45167/SPS_2017_01.pdf?sequence=1&isAllowed=y
- Vandierendonck, C. (2014). Public spending reviews: Design, conduct, implementation. Brussels, Belgium: European Commission.
- World Bank. (2004). Monitoring and evaluation: Some tools, methods and approaches. Washington, DC: World Bank.

AUTHOR INFORMATION

Marc Robinson is a Swiss-based senior consultant specializing in public financial management and fiscal policy. His extensive publications on budgeting issues include Performance Budgeting: Linking Funding to Results, published by IMF/Palgrave Macmillan, and "Spending Reviews," in OECD Journal on Budgeting. Marc has advised over 30 countries, ranging from OECD countries to developing countries, on budgeting reforms. Earlier in his career, he was a senior civil servant, a professor of economics, and a staff economist at the International Monetary Fund (IMF). He is a member of the OECD's Advisory Panel on Budgeting and Public Expenditures and the IMF Panel of Fiscal Experts. His website is www.pfmresults.com.