

Linking Ongoing Performance Measurement and Program Evaluation in the Canadian Federal Government

Neil Maxwell

Office of the Auditor General of Canada*

ABSTRACT

La relation qui existe entre les systèmes continus de mesure du rendement et l'évaluation des programmes est telle qu'elle produit des systèmes qui sont aussi bien complémentaires que mutuellement avantageux. La mesure du rendement peut libérer le travail d'évaluation des questions opérationnelles et favoriser ces évaluations en fournissant un véhicule pour la collecte continue des données. L'évaluation peut être utile en faisant transférer la méthodologie et préciser les questions examinées au cours de la mesure du rendement. Cette complémentarité laisse entendre que les deux devraient être planifiés en même temps dans le cadre d'une stratégie commune.

Introduction

Not much has been written in Canada on the subject of the relationship between ongoing performance measurement and program evaluation. This is not the case in the United States, where the two are often treated together. For example, when the Evaluation Research Society published its *Standards for Evaluation Practice*, ongoing performance measurement was included as one of the evaluation activities covered by the standards (Evaluation Research Society, 1982). Acceptance in the United States of performance measurement as a form of evaluative activity owes much to the work of the Washington-based Urban Institute, the single most prolific source of discussion on the subject.

Performance measurement shares with program evaluation the root definition of being the systematic gathering, analysis and dissemination of verifiable evidence of a program's results. But the two differ in two key characteristics.

The most obvious difference is in the frequency of reporting. Performance measurement systems are designed to report results regularly, reflecting the cycle of program activities. Monthly, quarterly or annual intervals are typical. Because of this frequency, they usually form part of management's various manual and automated systems. Program evaluation is less frequent. In the Canadian federal government, as prescribed by the Office of the Comptroller General, each program is to be evaluated every three to seven years (OCC, 1981).

The second difference is that program evaluation asks broader questions about a program than performance measurement does. Both begin as a systematic gathering, analysis and dissemination of verifiable evidence of a

* The views expressed here should not be construed as the policy or position of the Office of the Auditor General.

program's results. Beyond this, program evaluation is intended to question the very existence of a program, whether the need for it continues, or whether other programs could meet that need better. This broader role has been given considerable prominence by the Office of the Comptroller General, covering two of their four generic evaluation issues – program rationale and program alternatives (OCG, 1981).

The separate treatment of program evaluation and performance measurement in Canadian literature may be partly explained by the different and separate evolution of the two functions in the federal government. They are governed by different Treasury Board directives, overseen by different central agency groups, and are almost always organizationally separated in departments. Few if any departments have placed responsibility for program evaluation and for designing and operating performance measurement systems together.

Perhaps even more important, they have tended to be treated as mutually exclusive. In the federal government this has, over time, taken on the appearance of a counter-cyclical trend, emphasis waxing on one as it wanes on the other. Since the early seventies, there have been two or three distinct periods in the relationship between the two functions. First, through much of the seventies, performance measurement was the centre of attention. Program evaluation was talked about, but little was actually conducted (Treasury Board Secretariat, 1974). In the late seventies and early eighties, after the creation of the Office of the Comptroller General, the balance swung 180 degrees, with impressive advances in the acceptance and use of program evaluation (Office of the Auditor General, 1983). Performance measurement was largely ignored, and the systems that had been in place in departments fell into disuse. We may now be in a third period, distinct from the second because of renewed interest in performance measurement systems as part of Treasury Board's Operational Planning Framework initiative, which requires the reporting of program results. The decision to focus on either program evaluation or performance measurement, but not both, has also been taken by the province of Ontario. It has concentrated on performance measurement through its Management-by-Results initiative (Management Board Secretariat, 1982).

The sundering of program evaluation and performance measurement has been unfortunate. That is the argument from which the principal conclusion of this paper will be drawn: that the two should be designed in concert to ensure that they are complementary forms of management information.

Is Ongoing Performance Measurement Necessary?

A question that is often asked is whether ongoing performance measurement is truly necessary. There are certainly those who feel that it is not. In the federal government this negative attitude is often based on memories of Operational Performance Measurement Systems (OPMS) which, like many of the other 'alphabet soup' management tools of the sixties and seventies, did not live up to expectations (McCrandell, 1975). Part of the difficulty was its limited applicability – the routine production processes for which it

was best suited constituted only a minority of public sector programs. As well, OPMS suffered from the reductionism common to many management information initiatives — the false hope that complex questions could be answered by measuring one or two indicators.

As these deficiencies become evident, the federal government attempted, by Treasury Board Directive 76-25, to broaden the concept of performance measurement to include the monitoring of effectiveness and to extend it to all government programs. Performance measurement was also given a role in supporting Program Forecasts, the departmental budgetary requests that were made annually to the Treasury Board.

None of these modifications were successful in overcoming the scepticism about performance measurement that Operational Performance Measurement Systems had created. Even today, when discussing performance measurement, many public service managers immediately refer to experiences with Operational Performance Measurement Systems in the mid seventies. Such people still have a tendency to see performance measurement as a fad whose tenure is as fleeting as was Planning, Programming and Budgeting Systems (PPBS), Zero Based Budgeting (ZBB) and Management-by-Objectives (MBO).

The essence of performance measurement, however, is not transitory. The notion that a program's information systems should be designed to provide regular information on the program's outputs and outcomes (or results) as well as its inputs, is surely axiomatic. It cannot be dismissed as simply a management system fad.

One argument sometimes put forward by managers is that program evaluation pre-empts the need for performance measurement. The following section will attempt to refute this by arguing that each function is important to the other's success.

How Performance Measurement Can Improve Evaluation

Where there are no performance measurement systems, evaluations often get clogged with issues that should be answered by ongoing measurement. Examples include issues such as client satisfaction, target group coverage, comparisons of efficiency and cost effectiveness among regions, homogeneity of program implementation, etc. With an ongoing measurement system addressing such issues, program evaluation can concentrate on areas where its characteristics of in-depth analysis and multiple lines of evidence are needed. In cases such as these, the issue domains of the two functions are separate.

Another possibility exists for partnership between the two — one which might be called the smoke-detector model. Performance measurement can provide the first cut at detecting potential problems, which can then be explored by program evaluation. In such a model, the issue domains of the two functions overlap to some extent, with evaluation taking a more in-depth and selective look at results that have also been examined by ongoing measurement.

Perhaps the most persuasive argument for the latter model rests with the need in evaluations for repeated measurement over time. The accumulated

data base which performance measurement generates permits the use of more rigorous evaluation designs by permitting longitudinal or time-series evaluation designs. It is probably unnecessary to restate the strength of longitudinal designs. However, even where longitudinal designs are not chosen, performance measurement data bases can permit cross-sectional analysis between regions or districts (Poister, 1982), or simply provide data that, in the absence of performance measurement systems, may not be obtainable within the time and resources of an individual evaluation. To illustrate, the Auditor General of Canada (1983) in an audit of 19 government departments and agencies found that:

Almost 80% of the (sample of 43) studies relied to a significant degree on administrative data. In doing so, many of these studies encountered problems because these data were incomplete, inaccurate or poorly stored. Often data which would have been of great value in assessing program effectiveness and which could have been gathered at small marginal cost were not being collected. Such data often represent the only practical way to assess with any precision the impact that a program has had on its clients. Accordingly, these data bases should be given close attention well before they will be needed for program evaluation purposes. Additionally, if these data bases are well developed, the information in them will often be of great value to line managers responsible; for program operations. (Auditor General of Canada, 1983:121)

The Auditor General concluded that this constituted one of the most significant problems facing program evaluation.

How Evaluation Can Improve Performance Measurement

Having evaluated an issue permits evaluators to refine subsequent ongoing measurement that will take place before the next program evaluation. In this idea lies the full power of the concept of cyclical evaluation (evaluating every program every three to seven years) that is part of the Office of the Comptroller General's evaluation policy. There are many examples of evaluations that should have led to an ongoing measurement system. For example, an evaluation several years ago of a government information service found that significant parts of its target clientele were not using its services. However, no ongoing monitoring of client characteristics was undertaken subsequent to the evaluation, so that if there continues to be no significant improvement, this information will not be known until the next evaluation.

Evaluations also develop methodology that can be transferred to performance measurement systems. In one recent example, a department decided that the front end investment in questionnaire design, pre-testing, sample frame development, sampling, and so on was so large that the cost of ongoing monitoring of client satisfaction would not be warranted. However, program evaluators in the department had completed a large client satisfaction survey which could be easily adopted to provide time-series and even longitudinal information on client satisfaction.

Perhaps the most important area where program evaluation can improve

performance measurement in the short term, is in the transfer of system design techniques. One of the greatest problems facing the designers of performance measurement systems is how to go about designing such a system. As departments dust off the performance measurement concept to fulfil their Operational Planning Framework responsibilities, a new generation of project leaders has had to begin from first principles and reinvent and relearn the lessons of those who worked on performance measurement in the 1970s. Not surprisingly, many of these projects are misdirected. The most serious problem is the inordinate time that is spent developing and redeveloping indicators, leaving little time for fundamentals such as identifying the generic parts of the program, effectiveness and efficiency issues of interest to management, etc.

While project leaders are facing this problem, the very skills they need exist in the program evaluation functions located in their own departments. Designing a performance measurement system involves tasks that are identical to those required for a program evaluation assessment. Ironically, in many if not most departments, program evaluators have been involved in Operational Planning Framework projects, but only in questions of clarifying objectives, or ensuring that the component structure (the departmental structure for evaluation purposes) is consistent with the planning element structure (the departmental structure for management purposes).

Planning Evaluations and Performance Measurement Systems

One solution — a solution which ensures that evaluation and performance measurement will have complementary roles — lies in broadening the use of evaluation assessments. Making evaluation assessments a routine practice has been one of the most important aspects of the Office of the Comptroller General's evaluation policy, but their usefulness has been limited because they are followed only by program evaluations. Joseph Wholey (1979) proposed a broader use of assessments. An assessment might lead to administrative monitoring, performance monitoring and/or to intensive evaluation.

The idea is still timely. The steps of an evaluation assessment which design an evaluation can also serve to design a performance measurement system. These are: defining the scope of the program units to be measured (the component structure); identifying the key activities, outputs and outcomes and the linkages tying these together; identifying the results issues of interest to management; defining measures to address each issue; designing data collection methods to collect each measure; planning the analysis that will be undertaken to answer each issue; costing and packaging different measurement options; and obtaining agreement from the client(s) as to which option will be implemented. A combined management information strategy, involving regular monitoring and more periodic in-depth analysis of issues, would be the logical outcome of an evaluation assessment.

According to the Office of the Comptroller General's policy, frameworks are to be completed for new and renewed programs. They are intended to clarify objectives and to provide a plan for a future evaluation, including

collection of data in the interim. In 1983 the Auditor General found that only 6 frameworks had actually been completed in 19 government departments and agencies (Auditor General of Canada, 1983). Review of recent OCG data suggests that more departments are now undertaking frameworks. However, this is generally only for new programs, which form only a small part of the range of programs that could benefit from a plan for ongoing data collection. Moreover, in my experience, most have focused on collecting ongoing data solely for the use of future evaluations, not for the use of program management.

Conclusions

The separate treatment of performance measurement and program evaluation in Canada, particularly in the federal government, is an unfortunate sundering. The two systems, when designed in concert, contribute to the success of each other. Performance measurement can either free evaluations from operational issues or can improve evaluation rigour with repeated measurement. Evaluation can help refine the issues examined by performance measurement, and can transfer methodology to performance measurement. This complementariness suggests that the two should be planned together as a combined information strategy.

Program evaluation in the Canadian federal government has made dramatic advances in the past five years. After several swings of the pendulum toward and away from ongoing measurement, we appear to be at a point where attention is again turning to it. Hopefully, program evaluation is now sufficiently well entrenched that considerable attention can be given to making evaluation and performance measurement complementary systems.

References

- Auditor General of Canada (1983) *Report of the Auditor General of Canada for the Fiscal Year ending 31 March 1983* Ottawa: Government of Canada Chapter 3
- Comptroller General of Canada (1981) *Guide on the Program Evaluation Function* Ottawa: Government of Canada
- Evaluation Research Society (1982) *Standards of Evaluation Practice* San Francisco: Jossey-Bass
- Management Board Secretariat (1982) *Managers Guidelines to Managing by Results* Toronto: Government of Ontario
- McCrandell, J.Q. (1975) "Some reflections on administrative reforms in the federal government" *Optimum* vol.6, no.2
- Poister, T.H. (1982) "Performance Monitoring in the Evaluation Process" *Evaluation Review* vol.6, no.5
- Treasury Board Secretariat (1974) *Departmental Planning and Evaluation Groups in the Federal Government* Ottawa: Government of Canada
- Wholey, J.S. (1979) *Evaluation: Promise and Performance* Washington: The Urban Institute

Comments by John Mayne on the Paper by Neil Maxwell

“The Supply Side Lives On”

While I knew Neil Maxwell was writing a companion paper, and indeed heard an early version of it at **Evaluation '85** in Toronto, I did not read his paper until after mine was submitted. And yet, while we do agree on some things, I could not have asked for a better foil. Maxwell clearly believes that information is inherently good and that information systems therefore “... should be designed to provide information on the outputs and outcomes of a program ...” presumably for most programs. The only evidence the paper offers to support this is the statement that “This is surely axiomatic”!

In my paper, I suggest that it is exactly this supply-driven perspective that dooms most attempts at developing useful information systems. There are reasons why past attempts by sensible and well-meaning persons have not proven very successful, but Maxwell eschews an appraisal of this history for a plea for (seeming) rationality. Furthermore, by using the term “results” to cover both outputs and outcomes, Maxwell fails to distinguish the important differences and relative strengths between ongoing performance and program evaluation systems. As a result, the demand for management information is essentially ignored and little insight is offered on what is likely to make ongoing program performance information systems successful. Thus I have a fundamental problem with Maxwell's paper: I don't think pleas for more information really gets us anywhere.

On several technical points we agree. As my paper suggests but does not belabour, there is room for forging a better link between ongoing and periodic information systems. However, I would not suggest that a typical product of an evaluation assessment in the federal government should be recommendations for an ongoing information system rather than an evaluation study, although such recommendations might usefully be in addition to the evaluation study. One should not confuse an evaluability assessment with an evaluation assessment. To not carry out an evaluation study after an evaluation assessment means that no additional information could usefully be collected which would reduce uncertainty about the program and its results. This should be a rare event, and certainly would require more than, for example, unclear objectives.

On the other hand, as Maxwell suggests, evaluation frameworks would seem to be well-suited to recommending ongoing program performance information systems. Such recommendations could be very useful by-products and be used by evaluators to assist program managers with their day to day management — as opposed to senior managers, the principle clients for evaluation. Indeed, evaluation frameworks are in effect accountability frameworks for program performance and hence could be seen as cornerstones of an accountability regime. As such, they should be developed for all programs where accountability is an issue and not only for new and revised programs, as is current policy in the federal government.

Comments by Neil Maxwell On The Paper by John Mayne

“Information Demand Is The Name Of The Game”

Mr. Mayne's paper and mine do make good foils, but not for the reasons he claims. Mr. Mayne incorrectly ascribes to me the view that supply is the limiting factor in the success of information systems. His evidence for this is my claim that information systems “should be designed to provide information on the outputs and outcomes of a program”.

Nevertheless, the argument I was making – that ongoing performance measurement is useful and necessary – is fully accepted in Mr. Mayne's paper. He states “. . . this (operational) information is needed. Program managers need information on outputs in order to manage better and make operational decisions . . .”. He needs to look no further than his own work to find a foil.

In fact Mr. Mayne and I share the view that information demand has been given insufficient attention by many system architects. His criticisms of the Operational Planning Framework exercise are particularly timely as are his observations on the inappropriateness of systems maintained solely for external resourcing uses. I have long felt, as he does, that the key to successful performance measurement systems is to integrate them in the budgeting and work planning cycle. I hasten to add, however, that while this works to ‘push-start’ a system, individual managers must also find the information useful in their own areas of responsibility, not only in demonstrating results to their superiors.

Where we differ is in the question of whether the issue domains of program evaluation and performance measurement should be separate or can usefully overlap. Mr. Mayne correctly points out that the characteristics of the two differ markedly. I believe they can still contribute considerably to each other's success. He appears to reject the notion that evaluation can usefully provide a periodic in-depth analysis of issues also monitored by performance measurement systems. Thus, he relegates the two to quite separate lives. Unfortunately he does little, apart from brief remarks in his accompanying comments, to suggest how the two might be managed for their mutual benefit or to minimize overlap. In fairness, an author cannot cover everything in a single paper. Nevertheless, it is unfortunate that he tends to reinforce the notion that the two are entirely separate functions and that there is little need to co-ordinate the two.