Universalia Occasional Paper
No. 40, March 2000

Program Evaluation in the Year 2000

Gary Anderson[1]

This paper reviews the origins of program evaluation and its role in public sector management accountability. The paper considers some of the changes in the agenda over the last half century. Finally, three contemporary issues facing public sector evaluation are explored.

Origins of Program Evaluation

The early manifestations of program evaluation include evaluation research in such fields as education, agriculture and management.

In education, the origins of program evaluation go back more than a century. In the 1840s, after two hundred years of Puritan influence on education, the American Common School came under scrutiny. In the mid-1840s a conflict between the Boston schoolmasters and the State Board of Education led to the notion of measurement on the effectiveness of schools and in 1845 the first systematic school survey, using printed tests, was conducted there. The results of the survey were reported in tables indicating percent of correct answers and so forth. The focus was on program effectiveness, though it also set the foundations for standardized testing – an idea that became more feasible after 1870 and the invention of the steel pen. Subsequent studies in education were both research-oriented and investigative such as the study led by Abraham Flexner in 1918 of the innovative schools of Gary, Indiana or the Eight Year Study of progressive education in the 1930s led by Ralph Tyler whose influence became dominant in evaluation for the next half century.

In terms of perspectives and methodologies, agriculture played a considerable role. Much of the research in statistics early this century derived from agricultural research that used experimental methods to maximize the economic returns of crops. It relied on random assignment of treatments to plots and influenced the practice and even the vocabulary applied to evaluation research. Thus, evaluation researchers tried hard, if unsuccessfully, for many decades to find the ideal equivalent control group. The dominant paradigm was the experiment in which different program variations could be easily compared. Formalization of quasi-experimental methods in the early 1960s allowed evaluation researchers to statistically remove group inequalities that could not be successfully accommodated by matching or random assignment. However, all of these approaches relied on quantitative measurement and were limited to questions that lent themselves to quantitative measurement.

Another quantitative American influence was the field of scientific management. By the early part of this century, U.S. industry was highly developed and demands for increased productivity gave rise to studies of the industrial enterprise. Frederic Taylor (1856 -

[1] The author is President of Universalia Management Group, a Canadian management consulting firm engaged in program evaluation. He has written and conducted research on evaluation methodology including questionnaire development, interviewing, and conducting focus groups. His popular book *Fundamentals of Educational Research, 2nd edition* (1998) is acknowledged as a leading source of information on applied methodology. He is also co-author of *Institutional Assessment: A Framework for Strengthening Organizational Capacity for IDRC's Research Partners* (1995) and *Enhancing Organizational Performance: A Toolbox for Self-Assessment* (IDRC 1999).

1915) initiated the field of scientific management through his tremendous success as an industrialist. Time and motion studies as well as variations in working conditions in factories that led to discovery of the famous “Hawthorne” effect, were typical. These experimental approaches were similar to those in agriculture, and as was the case for studies in that field, the emphasis was almost exclusively on measuring "how much" rather than understanding underlying qualitative differences.

Not all the effort was strictly quantitative. From the earliest days, critics attacked the bean-counters, and alternative paradigms were advocated. Thus, the quantitative doctrine of Ralph Tyler was challenged by anthropologists and ethnographers intent on understanding the meaning of underlying events and symbolic interactions.

In conclusion, a variety of disciplines have adapted social science methodology to the field of program evaluation, providing credible and useful information for decision makers and program stakeholders (e.g. Cook & Campbell, 1979; Cronbach et al., 1981; Patton, 1986). Thus, by the late 1970s, American evaluation supported several scholarly journals, an annual output of dozens of evaluation monographs and two national professional societies.

Evaluation in the Public Sector

By the early 1950s, and with the rise of the World Bank as a development institution, people became increasingly interested in the concept of maximising social and economic benefits. Evaluation was one essential tool. In the 1960s and 1970s, many countries adopted the view that government needed disciplined and effective systems of accountability, and policy making, planning and budgeting (Adie & Thomas, 1987). Evaluation was seen as an essential ingredient of both. Through the introduction of PPBS to the U.S. Defense Department, Robert McNamara attempted to introduce private sector management techniques to the public sector. He later brought these ideas to the World Bank.

At the same time, people began to feel that public sector policy making, planning, and budgeting needed the discipline of systems such as MBO (management by objectives), PPB (planning, programming and budgeting), OPMS (operational performance measurement systems) and various forms of benefit/cost analysis (Hartle, 1973). This view originated in efforts to rationalise planning and resource allocation in the United States Department of Defense (Chelimsky, 1985). Evaluation was seen as a key element of the policy making, planning, and budgeting cycle. In the light of experience, many enthusiastic proponents of the more extreme forms of systematic rational planning and budgeting and of performance measurement moderated their views (e.g. Hartle, 1976; Reid, 1979). Critics asserted that performance measurement systems may be more appropriate in the private sector, where outcomes can usually be expressed in terms of dollars or other countable units.

Most developed countries established evaluation offices under such labels as the Office of the Auditor General. As well, evaluations were carried out sporadically by planning and evaluation units in individual government departments. During this period in Canada, the Auditor General reported that few successful evaluations had been conducted, and he made repeated calls for more and better evaluation. In 1977, the Canadian Treasury Board adopted a new policy: “Departments and agencies of the federal government will periodically review their programs to evaluate their effectiveness in meeting their objectives and the efficiency with which they are being administered (Treasury Board of Canada, 1977)”. Evaluations were seen as consisting of both:

- a *disciplined study* of evaluation issues, which provides senior management with relevant, credible, objective, and timely findings, conclusions, and recommendations; and,

- a *review by senior management,* using evaluation and other information, which results in program decision making and appropriate accountability reporting.

In other words, an evaluation within a federal department or agency was seen as consisting of both the process of evaluating certain issues and the deliberate use of that evaluative information by management.

The Canadian policy was further refined in 1981 to focus on outcomes, impacts and program alternatives, and by 1983, federal evaluation had been sufficiently established for the Auditor General to devote a chapter of his Annual Report to Parliament to it. He reported that: "...very real progress has been made. Most of the major departments and many of the agencies have the basic infrastructure for evaluation in place. Many of these are actively attempting to carry out evaluations. However, the quality of the evaluations needs to be improved." In addition, he reported that the central agencies did not use the evaluations produced and seemed little interested (Auditor General of Canada, 1983; Soper, Peterson, Maxwell & Shaffron, 1984; McRoberts & Hudson, 1985). A follow-up audit by the Office of the Auditor General in 1986 reported significant improvement in the quality of evaluations (Auditor General of Canada, 1986).

By 1985, in Canada, there was general agreement within the federal evaluation community on a number of lessons learned. (Ulrich, 1984; Office of the Controller General, 1985).

- Evaluation is only useful when management is concerned with program results.
- Evaluation should be integrated structurally into ongoing strategic management and accountability processes.
- Evaluation needs skilled and sensitive practitioners who can put the discipline of evaluation to work in specific contexts.
- Most of the information needed for strategic management cannot be derived solely from ongoing MIS, but needs to be gathered and analysed in periodic studies (evaluations).
- Evaluation can have considerable impact on policy development and program design if effective communication exists among the various groups concerned and if line managers and program clients are involved substantively in the evaluation process.
- Cabinet committees and central agencies can be effective users of evaluation findings.

Despite considerable success in establishing departmental evaluation, observers continued to consider it fragile and speculate whether it would play a substantive role in federal decision making and accountability (e.g. Doern & Phidd, 1983; Mayne & Mayne, 1983; Rayner, 1986).

By the late 1980s, the mood in developed countries had changed with conservative forces assuming power. Margaret Thatcher in Britain and Ronald Reagan in the US adopted policies to provide more public accountability. This administrative and governance revolution was termed the New Public Management (NPM) and it has swept across most Western democracies. NPM represented a withdrawal from centralised and control-oriented systems of governance towards a more "entrepreneurial" style of governance representing decentralisation and empowerment, as well as significant cost-cutting.

Unfortunately, implementation of these policies focussed almost entirely on study of efficiency as the dominant evaluation issue. Public policy was to cut government deficits and to reduce the size of government. This translated into using evaluation to focus on cost-effectiveness rather than more inclusive evaluation of success in achieving social goals. Thus, evaluation had moved from a means of making decisions about the allocation of resources across programs to a tool for reducing costs within a program. In practical terms, auditing became a viable alternative to evaluation with its focus on standards and best practices (See Canadian Comprehensive Auditing Foundation, 1987). Indeed, the new public management has tended to reduce the visible presence of evaluation in most countries.

Critical Issues

At this time, there are many critical issues facing the field of program evaluation. Three such issues are presented here.

Dealing Creatively with Measurement Paranoia

There is a tendency to value only what can be easily measured. Measurement is in support of the rational model for decision making and accountability. As Sir Josiah Stamp is said to have observed:

> The Government is very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he pleases.(cited by Smith, 1989)

In most policy areas and in many programs, conclusive answers to evaluation questions are hard, if not impossible, to achieve. W. Edward Deming, the father of Total Quality Management and Statistical Process Control, is frequently misunderstood as a supporter of such measurement techniques. In fact, he was strongly opposed to such practices, stating for example: “Focus on outcome (management by numbers, MBO, work standards, met specifications, zero defects, appraisal of performance) must be abolished, leadership put in its place (Deming, 1986).” Handy (1996) says that traditional measurement systems, “conditioned by the philosophy of the audit . . . to count only what they can put their finger on and cross their hearts about” results in more attention to efficiency than effectiveness and to lack of attention to what is really important, even if difficult to quantify, such as a business’s intellectual assets and how it is positioned to respond to the future.

For these reasons, the private sector is moving away from a narrow measurement focus on outcomes towards a broader approach (e.g. the “balanced scorecard” of Kaplan and Norton, 1992) which looks at a wide range of measures, including difficult-to-measure factors such as focus on innovation and learning. Also, at the organisational level, Universalia has developed a balanced approach (Lusthaus, Adrien, Anderson and Carden, 1999). Recent evidence from the United Kingdom (Ruddle and Feeny, 1997) indicates that companies that focus their measurement systems on traditional short-term financial targets are less effective than those which look at broader, hard-to-measure issues such as employee behaviour, skills and infrastructure. In other words, organisations that focus on quantifiable measures of outcome are less effective at achieving their desired outcomes than those that take a broader perspective!

Evaluators must be clear and up-front about such limitations, but they must continue to advocate the value of the disciplined search for insight, of the challenge to the conventional, and of well-thought-out program innovations. If what is measured is what is valued, evaluators need also to introduce more creative ways of measuring things that are important even when they are difficult to measure. A concept such as national pride may be elusive to measure but may be an important outcome of some programs.

Avoiding a Singular Focus on Efficiency

Program evaluation may have significant potential for providing information to support resource allocation decisions, to help determine the value obtained from tax dollars and help public servants manage for results and take responsibility for results, but not if the concern is limited to maximising efficiency. This tendency is reflected in some countries by a new language where efficiency is couched in such terms as performance review. From time to time, systematic disciplined studies are needed which examine all dimensions of the program – that is, comprehensive program evaluations focussing also on long term impact.

There is increasing evidence that current approaches to performance review are repeating the mistakes of the past. Mintzberg (1994), in his influential book *The Rise and Fall of Strategic Planning*, has pointed to the failure of efficiency-based systems in Canada and elsewhere in both the public and private sectors. He indicates that these and similar measurement-based approaches have not resulted in improved performance or a focus on outcomes. On the contrary, they have hampered the development and implementation of effective strategy. The infamous Vietnam body counts, which used counts of "enemy" dead as an indicator of progress in fighting the war, still remain as the classic example illustrating how a performance management approach can result in horrible distortions. As Mintzberg (1996) says: "How many times do we have to come back to this one until we give up?"

As early as the mid-1980s, Canadian critics have charged evaluation with its generally self-serving nature and its failure to adequately address the issue of program rationale and relevance and the fiscal totality (the full range of programming, particularly special tax measures, regulation, and the combined benefits received by clients from different programs). These criticisms were agreed to be largely well founded, and they provoked a useful debate among federal evaluators and managers concerning ways of improving the credibility and utility of evaluation for a wider range of program stakeholders (Foote, 1986; Mayne, 1986; Rutman, 1986; Ulrich, 1987).

In Canada, the Treasury Board has adopted three core evaluation issues. These issues are meant to be considered in the planning of all evaluations. The first is relevance. This speaks to the question of whether the program is consistent with existing government and departmental priorities, and whether it continues to address an actual need. The second is the success of a program in meeting objectives established for it. The third is the cost effectiveness of the program. That is, whether means used by the program to achieve its objective are appropriate and efficient compared to alternatives.

The Implementation Challenge: Using Evaluation Results to Improve Program Performance

The ultimate objective of program evaluation is to improve programs. However, doing so requires sound evaluations and a commitment by responsible organisations for implementation of their recommendations.

The present context is a difficult one in this respect. For example, in Canada we find a changing environment in which various factors, such as the attempt to place a greater focus on results-based management and monitoring, have changed how evaluation links into the budget management process (Office of the Comptroller General, 1991). The government has conducted a major restructuring. It reduced the number of departments, and it significantly restructured a number of the remaining ones. This created enormous problems and challenges for planning and conducting evaluations because evaluations were suddenly aiming at moving targets. It also made it difficult to identify who was accountable for what programs and what activities. Thus, one could not link evaluation to program improvement because the program always seemed to be moving.

An effective system of ensuring implementation of evaluation results requires incentives and rewards for managers who use evaluation to improve their programs. Only then can we avoid the current tendency of managers to attempt to control the evaluation findings and to block recommendations that require them to change the way they manage.

References

Adie R. F, & Thomas P. O. (1987). *Canadian public administration: Problematical perspectives* (second edition). Toronto: Prentice-Hall.

Anderson, Gary. (1998). *Fundamentals of educational research.2nd Edition.* London: Falmer Press,
276 pp.

Auditor General of Canada. (1983, 1986). Annual Report of the Auditor General of Canada to the House of Commons.

Canadian Comprehensive Auditing Foundation. (1987). *Effectiveness reporting and auditing in the public sector.* Ottawa.

Chelimsky, E. (Ed.). *(1985). Program evaluation: Patterns and directions.* Washington: American Society for Public Administration.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cronbach, L., et al. (1981). *Toward reform in program evaluation.* San Francisco: JosseyBass.

Deming, W. Edward. (1986). *Out of the crisis*. Cambridge: Massachusetts Institute of Technology Center for Advanced Engineering Study.

Doern, O. B., & Phidd, R. W. (1983). *Canadian public policy: Ideas structure process.* Toronto: Methuen.

Foote, R. U (1986). The case for a centralised program evaluation function within the government of Canada. *The Canadian Journal of Program Evaluation, 1(2),* 89—96.

Handy, Charles. (1996). *Beyond certainty: The changing worlds of organizations*. Boston: Harvard Business School Press.

Hartle, D. O. (1973). A proposed system of program and policy evaluation. *Canadian Public Administration, 16(2),* 243-266.

Hartle, D. O. (1976). Techniques and processes of administration. *Canadian Public Administration, 19(1),* 21—33.

Kaplan, Robert S. & Norton, David P. (1992, Jan-Feb). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 71-79.

Lusthaus, Charles, Adrien, M-H., Anderson, G., & Carden, F. (1999). Enhancing organizational performance: A toolbox for self-assessment. Ottawa: IDRC.

Mayne, 1. (1986). In defense of program evaluation. *The Canadian Journal of Program Evaluation,* 1(2), 97—102.

Mayne, *I.,* & Mayne, R.. (1983). Will program evaluation be used in formulating policy? In M. Atkinson & M. Chandler (Us.), *The politics of Canadian public policy* (pp. 267—284). Toronto: University of Toronto Press.

Mintzberg, Henry. (1994). *The rise and fall of strategic planning*. New York: The Free Press.

Mintzberg, Henry. (1996, May-June). Managing government, governing management. *Harvard Business Review*, 75-83.

McRoberts, H. A., & Hudson, 1. *(1985).* Auditing program evaluations: The Canadian case. *Accounting Organization and Society 10(4),* 493—502.

Office of the Comptroller General. (1985). *Background notes for a seminar on program evaluation* (March 7—8). Montreal.

Office of the Comptroller General. (1991). *A compendium of evaluation utilization.*

Patton, M. O. (1986). *Utilization-focused evaluation* (2nd ed.). Newbury Part Sage.

Rayner, M. (1986). Using evaluation in the federal government. *The Canadian Journal of Program Evaluation, 1(1),* 1—10.

Reid, I E. (1979). The failure of PPBS: Real incentives for the 1980s (The past is not prologue). *Optimum, 10(4),* 23-36.

Ruddle, Keith & Feeny, David. (1997). *Transforming the organisation: new approaches to management, measurement and leadership.* Oxford: Templeton College Information Centre and Library.

Rutman, L (1986). Some thoughts on federal level evaluation. *The Canadian Journal of Program Evaluation, 1(1),* 19—28.

Smith, M. F. (1989). *Evaluability assessment: A practical approach*. Boston: Kluwer.

Soper, N., Peterson, K., Maxwell, N., & Shaffron, R. (1984). Facilitating use: The Canadian experience. *Optimum, 15(4),* 6441.

Ulrich, M. (1984). Rigour versus relevance in program evaluation. *Optimum, 15(4),* 56—63.

Ulrich, M. (1987). *Program evaluation in the federal government.* Based on presentation at the National Annual conference of the Canadian Evaluation Society.

Universalia is a Management Consulting Firm located at:

5252 De Maisonneuve West, Suite 310
Montréal, Québec, Canada H4A 3S5
Tel: 514-485-3565, Fax: 514-485-3210
Website: www.universalia.com

\\univers1\sys\!umg\!corporate\anniversary\prog_eval\2000_ghana.doc