New and creative approaches to program evaluation are required for evaluation to play a meaningful role in the public policy process. The unique and complex relationship between politics and program evaluation presents a diverse range of challenges that cannot be adequately addressed by traditional scientific approaches to evaluation. First, I will focus on the potential contribution of program evaluation to the public policy process which will be compared against its current actual role in political decision making. Then, based on a review of the literature, I will advance the argument that creative qualitative approaches to evaluation are requisite to improving the relevance and utilization of program evaluation in a political operating environment. Finally, I will offer specific suggestions for change to the process of program evaluation, placing particular emphasis on the need for enhanced stakeholder participation in evaluation.

Potential contribution of program evaluation to political decision making

Program evaluation should supply definitive data for planning purposes, causal knowledge as to which program variables to change to achieve desired outcomes and cost-benefit information towards selection of the intervention with the greatest payback. According to Weiss, there is “much hoopla about the rationality that [evaluation] would bring to the untidy world of government.”

However, the effects of evaluative research on shaping public policy have been relatively limited. On the whole, it would appear that few studies have had a notable impact on political policy making. Numerous authors confirm that decisions respecting the termination or reduction of ineffective public programs based on evaluative findings are virtually non-existent. Canadian practice appears to have confined program evaluation to “more of a management tool than a real challenge to government programs.”

A major paradigm shift in practice is required if program evaluation is to assume a meaningful role in the public policy process.

Judy Layne is the Dean of Science and Health at the Saskatchewan Institute of Applied Science and Technology. She is currently completing the Master of Public Administration Program and is a double-gold medalist in her undergraduate studies in science, education and personnel management. At the present time, her areas of academic interest include program evaluation and performance measurement in the public sector, and alternative service delivery models such as public-private sector partnerships. Ms. Layne has worked in both the public and private sectors in human resources management, and she has published articles on various human resource management and general management topics in professional journals.

* This article is an abridged version of the paper that Judy Layne wrote as part of her Master of Public Administration Program. Her paper was designated the best 1998 graduate student paper in the nation by the Canadian Association of Public Programs of Canada, and by the Institute of Public Administration of Canada – Manitoba Division.
Further, the publication of evaluative findings does not appear to be widespread, with reports usually retained in-house for use by government departments. A number of authors note that evaluation practice in Canada is not designed to meet the information needs of the citizenry, but rather of program managers. There are naturally some notable exceptions, including the 21-volume Nielsen Task Force report that is a matter of public record. In general, however, evaluation findings are released to the press and/or the public only after extensive internal review and editing.

Moreover, while there is general consensus in the literature that stakeholder involvement is an integral component of the evaluation process, such involvement is clearly not being pursued. Schelling, as quoted by Greenberger, feels that “analyzing the interests and the participants may be as important as analyzing the issues” in evaluation. The President of Treasury Board emphasizes that government “must work to meet public expectations.” Yet, government has no way of knowing if programs are meeting expectations unless that public is involved in the evaluative process. Renouf feels that “we have ... been deepening our capacity for ‘Ottawa to evaluate Ottawa’ rather than for the Canadian community to evaluate its federal government.”

Program evaluation and the public policy process

Whys and hows of evaluation

Program evaluation may be undertaken for a wide variety of reasons, generally falling under the umbrella goals of improving public policy and/or service delivery.

Pal states that the primary reason for undertaking evaluative studies is to determine if “policies, programs and operations are working well.” Pal’s statement of purpose lacks clarity, and other authors elaborate on what “working well” might entail. Rossi and Berk describe the objectives of evaluation broadly. They see evaluation being conducted to answer important questions that arise during policy formulation, program design and implementation, and to test the efficiency and effectiveness of existing and proposed programs. It is noteworthy that the references in the literature to evaluative studies being carried out on programs under consideration are rare.

Patton suggests that program evaluation can serve useful purposes over and above political decision making. Such objectives may include: fine tuning of program implementation, development of clearer program goals, procurement of more definitive information on client needs and stimulation of insightful thinking about a program (the outcome of which may not surface until some future point). Simeon believes that the mere undertaking of an evaluative study can create a window of opportunity for change, and cites the comments of a Nielsen Task Force Review interviewee who felt that this study broke blockages in the system and infused new ideas into government.

The purpose of evaluation may be purely political and may be directed at building the image of a program or of government itself. Patton suggests that politicians may use evaluation for the express purpose of killing a program, but no references are available to support this position. It is reasonable to speculate that political aims may also include justifying a previously arrived-at position.

Rossi, Freeman and Wright take a pessimistic view of the purpose underlying the conduct of program evaluation. These authors conclude that the sponsorship of an evaluative study is an implicit expression of doubt that the program may not be as effective as hoped.

The statement of purpose offered by Berk and Rossi perhaps best reflects the political context within which program evaluation must operate. These authors see evaluation as providing “the best information possible on the key policy questions within the given set of real-world constraints.” In my view, the “best possible” information is useful, relevant, and includes political considerations along with process and outcome measures.

Evaluation methodology may be categorized as either impact- or process-focused. Impact evaluations, referred to as outcome or summative studies, are conducted post-implementation. Such studies provide a retrospective assessment of program outcomes against objectives. While the majority of evaluative studies focus on program impact, this is often too late to have any meaningful role in shaping policy decisions, since as Palumbo aptly points out “policies take on lives of their own once they are implemented.”
Impact evaluation bears a causal knowledge problem in that it views the program in question as a causal variable and attempts to isolate its impact from that of other stimuli. Such causality is difficult, if not impossible, to establish. Cost-benefit studies, which fall under the umbrella of impact evaluations, carry a further problem of attempting to associate a dollar figure with the often intangible costs and benefits of public programs.

Process evaluations, referred to as formative studies, analyze program implementation. Such evaluations attempt to assess whether a program is being delivered to its intended audience in its intended “dosage” and to answer the question of how a program may be enhanced. Palumbo and Nachmias note that while evaluation is traditionally seen as providing summative information with which to make decisions on whether to continue or terminate public programs, there appears to be slowly increasing emphasis on process evaluations. Perhaps this change is based in part on a recognition that such “life and death” program decisions are rarely made in the political arena.

Comprehensive evaluations include both formative and summative components. Such blending of process and impact data seems logical and complementary, since the quality of program implementation, as measured through process evaluation, is not necessarily synonymous with the quality of the program itself, as measured through impact evaluation. It is my position that the potential contribution of program evaluation to political decision making can be maximized by providing a seamless continuum of evaluation findings through the integration of process and impact review.

**Program evaluation as scientific enterprise**

Program evaluation has evolved as a scientific enterprise. Indeed, it must be acknowledged that progress in program evaluation has been made possible by the solid foundation of scientific quantitative analysis upon which it is built.

Today, program evaluation continues to present itself as a scientific and systematic discipline. It has been varyingly described as a “science of information” and as “the dismal science of policy analysis.” The literature confirms that, notwithstanding some degree of change in practice, the dominant paradigm is still one of scientific inquiry. Much emphasis is still placed on methodological rigour in keeping with the positivist model of science. Precision in evaluation methodology is, in fact, often seen by those in the field as providing immunity to political and other contextual influences.

**Challenges of conducting program evaluation in a political operating environment**

The challenges of conducting program evaluation within a political context are numerous and varied – for example, the volatility of the operating environment in which evaluation is conducted. A climate characterized by ongoing downsizing and budget cuts ultimately means that fewer dollars are available for evaluative research, and that the priorities of those working in this climate are understandably placed on things other than evaluation. Pal aptly notes the paradox of a financial situation that underscores the need for evaluation yet at the same time restricts its practice. The high cost of many evaluative studies further complicates this issue. Derlien reports that 20 percent of evaluations conducted in the U.S. in 1984 cost in excess of $100,000.

The time clocks by which politics and program evaluation operate are often at odds. Pressure for speedy completion of evaluative studies may be exerted by the annual budgeting process, in which rapid funding decisions are frequently made on the basis of incomplete data. The compressed time frames within which politicians must impress the electorate may make it more politically prudent to launch a new program rather than undertake a review of an existing one.

Program management and staff tend to see evaluative reviews as threatening. This is understandable, since this group may be seen to have the most at stake in program continuance. As evaluative findings often deliver unpleasant truths, Pal notes a temptation to “shoot, or at least ignore, the messenger.” McCleary, as cited by Kelly,
cynically sees the response of program staff to evaluation as follows: where possible, workers use statistics not as statistics, but as instruments to reduce the workload. This criticism is perhaps too harsh, and a more realistic perspective is offered by Rogers who quotes Douglas Hartle as saying “It is the rare dog that will carry the stick with which it is going to be beaten.” Program administrators also feel pressured to present evidence of effective programs in view of political intolerance for failure. Given the apparent rarity of “positive” evaluations, we could expect to see a strategy of one program seeking inclusion with other programs in an evaluative study in order to look good comparatively.

Program objectives frequently offer poor yardsticks for measuring success. Such objectives are often non-specific, non-measurable, inflated and sometimes incompatible. Articulated objectives may reflect a hurried goal formulation process aimed at obtaining funding approval or a purposeful one aimed at garnering broader support. Rarely is thought given to developing comprehensive quantitative and qualitative indicators of effectiveness at the program development stage.

Summative assessments of program achievements are challenging for other reasons. First, the quantification of performance targets is more difficult in the public as compared to the private sector. It can in fact be argued that emphasis on quantifiable approaches biases program evaluation toward trivial notions that can be readily measured. Second, baseline data for post-implementation comparisons is often lacking. Third, as Rein points out, people’s view of desirability changes over time. Fourth, different stakeholders have different definitions of what constitutes success and may only be willing to accept certain measures of it. Finally, while scientific inquiry calls for keeping all variables constant, except the test variable, in order to isolate effects, the effects of public programs are difficult to isolate from other influences.

The process of program evaluation necessitates difficult trade-off decisions. Each evaluative situation is unique and generalizations are not possible as in the positivist scientific arena. An artful balance is required between the feasibility and cost of different evaluative methodologies and the benefits and utilization of evaluation findings.

The involvement of multiple stakeholders in evaluation is often seen as a threat to the integrity of study findings. Traditional scientific inquiry calls for researchers to maintain an objective distance from their subjects. If it is deemed important to incorporate multiple stakeholder positions in evaluation, an artful balance between consultation and neutrality must be maintained.

An interesting challenge is presented by the lack of set criteria by which to judge the relative worth of diverse stakeholder values. Guba and Lincoln express concern that different stakeholders’ values be treated equitably throughout the evaluation process. While it is noble to say that “all values are created equal,” I would argue that political decision making will ultimately call for prioritization or hierarchical ranking of values.

The need for a new approach to program evaluation

Webster’s Dictionary defines art as “the conscious use of skill and creative imagination.” An artful, more qualitative, approach to program evaluation is necessary to fulfilling its true potential in public policy making. This is not to suggest the negation of scientific methodologies in evaluative research, but rather to call for an added artful dimension of creativity in evaluation.

For purposes of this discussion, it is instructive to note the adage “art mirrors life.” Like art, evaluation research must reflect the contextual reality within which it operates. This reality may be characterized as follows. First, facts alone do not shape public policy; that is, evaluation findings must compete with a host of non-scientific factors that impact political decision making. Second, the process of political decision making itself is largely artful, consisting of discussion and negotiation to reach acceptable outcomes. The evaluation process arguably should follow suit. Third, as Patton points out, while the political world of special interests and multiple stakeholders is rational and logical, it is not the rationality or logic of economics. He goes on to say that the ideal of explicitly identifying goals and measuring their accomplishment is an exemplar of human reason and rationality. The reality of evaluation practice in a political context shows our shortcomings in trying to attain rational ideals.
Science and program evaluation search for different things and have different mandates. First, science searches for truth, while evaluation should search for information. As Qvortrup points out, evaluation must "go beyond systematizing empirical reality." Second, the mandate of science is to be objective, while program evaluation has no single objective or correct outcome. Third, while control is an underlying tenet of scientific inquiry, program evaluation cannot exert control over political realities. Rather, in Patton’s view, evaluation requires a problem-solving approach adaptable to change, as opposed to a technical approach that attempts to mold and define conditions to fit a preconceived model.

A more qualitative consultative approach is requisite to improving the relevance and utilization of program evaluation findings within a political operating environment. Brightman and Noble describe “decision scientists,” including traditional evaluation researchers, as “hopelessly naive and intellectually arrogant. Naive because they believe that problem solving begins and ends with analysis, and arrogant because they opt for mathematical rigor over results.” Greenberger notes that by focusing largely on technical and economic issues at the expense of questions of value to politicians, evaluators “put themselves at a disadvantage in the contest for attention and influence, and their work loses the impact it might otherwise have had.”

An artful approach is consistent with established professional evaluation standards of utility, feasibility, propriety and accuracy. While scientific methodologies such as sampling techniques are undoubtedly useful in program evaluation, on balance, conventional scientific criteria (validity, reliability, objectivity) are at odds with the norms and assumptions of the political arena.

Technical scientific evaluation processes cannot deal with value pluralism. Rather, artful methodologies must be employed to incorporate various value positions of multiple stakeholders. Like Katz, some authors feel that the issue of value pluralism can have no resolution through analysis. Rather than “throwing out the baby with the bath water,” preferred positions include those of Greenberger who states that “facts alone cannot resolve [a] controversy rooted in subjective values and ideological positions,” and of Rossi, Freeman and Wright, who concur that facts are often on the periphery of debate whose main themes center on values.

Kelly quotes Kimmel as adding support to the argument for new and qualitative approaches to program evaluation when he observes that “measurability is no index of importance.” Indeed, there are numerous examples of evaluation studies in which what has been omitted may be considered to be of equal or greater importance than that which has been included.

Finally, evaluation may be thought of as artful, since like art, one of its main benefits has been to provide people with a new perspective on issues.

Suggestions for change to program evaluation – carving out a meaningful role in the public policy process

Below I present specific suggestions for change to traditional scientific approaches to program evaluation so that evaluation can play a meaningful role in political decision making.

Adopt a situational problem-solving approach

The focus of evaluation must shift from one of methodological rigour to one of creative application. The work of several authors underscores the position that evaluative studies must take a situational problem-solving approach in which decision criteria are multiple and flexible. Creative applications should draw upon the entire range of methodological options available in the field and should employ the best combination of qualitative and quantitative techniques appropriate to a particular study.

Evaluative methodology should balance the goals of research validity with those of utility and relevance. I concur with the view of authors such as Patton who argue that utility should be the foremost criteria in making evaluative decisions. This means that the use of overly complex scientific methodologies that may threaten the relevance and usefulness of findings must be weighed against threats to validity. Concerns for technical precision must be examined in the context of meeting the spirit and challenge of professional evaluation standards.
**Increase the dissemination of study findings**

The dissemination of evaluation findings to Parliament and the public should be increased, thereby encouraging openness and scrutiny of public policy. The publication of evaluative results is viewed as integral to fostering a mindset of accountability within government. Various vehicles should be utilized for this purpose, including greater emphasis on evaluation in annual reports, increased media coverage, enhanced reporting to Parliament and better electronic access. Rist captures the importance of information sharing when he states that wider dissemination of evaluation results creates a climate for change distinct from that created by restricting findings to a need-to-know basis. While the President of Treasury Board agrees it is “important for taxpayers to have access to information about the cost of government activities and programs,” cost is not the only critical factor, and other indicators of program effectiveness are required. Should stakeholder participation be impracticable for any valid reason, the importance of disseminating the evaluation results as a means of fostering discussion and debate is seen to increase accordingly.

**Enhance stakeholder participation**

Stakeholder participation should be enhanced in all aspects of the evaluation process. Renouf cites the need for “reassertion of the community’s responsibility for its economic affairs.” Bemelmans-Videc and Conner point out that since government programs are aimed at steering social processes, causing changes in the actions of citizens, the “perspectives of stakeholders at the end of the causality chain should be represented.” The President of Treasury Board agrees that increased external participation would “increase confidence of central agencies in using departmental review findings for major decisions,” and the Comptroller General reasons that adding breadth and depth to existing forms of client consultation in public programming would create a receptive environment for change. Palumbo goes so far as to say that the views of program opponents should be incorporated in evaluative efforts. In my view, analyzing stakeholder needs and interests is equally as important as analyzing technical and economic issues. Furthermore, if evaluation is indeed a search for “the best possible information,” then the incorporation of stakeholder positions will provide politicians with complete data for informed decision making.

While I recognize that the effort, cost and complexity of stakeholder involvement must be addressed on a case-by-case basis, creative ways may be employed to deal with such concerns. Patton offers an example of using prospective jurors awaiting assignment as a focus group, by which inexpensive rapid data collection from a relatively large random sample of stakeholders was accomplished. It can be argued that program evaluation allows for two tiers of stakeholder participation, both of which would enhance existing methodology. The first tier sees stakeholders as interviewees, providing input during the implementation of evaluative studies. The second and preferred tier sees stakeholders as active participants in all aspects of evaluation design and implementation.

Considering program management and staff as stakeholders is worthy of special comment. The evaluation methodology must anticipate their understandable reactions to having programs reviewed and must take steps to address their reactions. Increased involvement in all aspects of the evaluative process is perhaps the best means of doing so. Potential concerns regarding program staff participation, namely, bias and subjectivity, should be weighed against the potentially greater benefits of their participation: increased commitment to the evaluation process, reduced cost of data collection and the development of in-house expertise.

**Redefine the evaluator’s role**

Consistent with a paradigm shift in the evaluation process, the evaluator’s role within that process must be reshaped. Guba and Lincoln offer a thought-provoking thesis on the evolution of program evaluation. These authors propose a “fourth generation” of evaluation in which traditional roles are retained, albeit in redefined states. In this model, evaluators serve as the primary study instrument, as opposed to more standardized measuring instruments. Rather than fixing an image through a scientific objective lens, evaluators “attempt to illuminate the scene” and promote understanding of multiple value
postures. Rather than acting as judges of data, evaluators become mediators of the judgmental process. In this role, the authors assert that decisions cannot be made until compromise is reached among varied value positions, or unless some players are significantly powerful to exert control over the process. As well, Guba and Lincoln acknowledge the latter as a common occurrence, but feel it is an immoral state of affairs. I would argue that these authors are perhaps too harsh in their criticism and somewhat naive in their pursuit of compromise. But I recognize that compromise is not always possible or necessarily the best decision. Instead, the availability of relevant information on different value positions can be seen to provide a more complete framework for political decision making.

**Balance the content of the evaluation’s findings**

Studies should identify unanticipated outcomes, even if negative, along with positive ones. While positive findings will obviously be of more political interest, unanticipated outcomes should be included if there is a true commitment to policy enhancement. Evaluation reports should focus on providing specific recommendations for action as opposed to offering critiques that are frequently academic, and they should incorporate program assets along with deficiencies to serve as a foundation for building commitment to the evaluative process.

**Align the evaluation process with the political decision-making process**

While evaluation findings are considered in the context of discussion and debate, the evaluation process itself should reflect discussion and debate among stakeholders. Nachmias observes that in most cases, neither evaluation findings nor the factors that induce policymakers to act the way they do are crystal clear. An evaluative process that provides for stakeholder involvement throughout could bring clarity to both questions.

**Expand the focus of program evaluation**

Finally, the focus and use of program evaluation should be expanded to that of a planning instrument. This suggestion represents a departure from the traditional role of evaluation as a retrospective measuring instrument. To be an effective planning instrument, evaluation must play a “life-cycle” role in program planning, implementation and outcome assessment.

**Conclusion**

A major paradigm shift in practice is required if program evaluation is to assume a meaningful role in the public policy process.

The dominant paradigm is still one of traditional scientific inquiry. At the same time, the interface between politics and evaluation presents complex challenges that traditional scientific approaches to evaluation fail to meet. This is why program evaluation currently falls short of its true potential as a political decision-making tool. Concerns regarding the relevance and utilization of evaluation findings can only be addressed through creative evaluation methodologies that use a situational problem-solving approach and a combination of qualitative and quantitative techniques.

Program evaluation’s contribution within the political arena may be enhanced through specific changes in approach, including: emphasizing stakeholder participation, increasing the dissemination of study findings, redefining the evaluator’s role and expanding the focus of program evaluation to that of a planning instrument.

A shift from rigour to relevancy in evaluative practice poses a serious challenge to the existing scientific paradigm and to the current balance of political power. Robert Henri’s 1923 description of an artist (as quoted by Patton) may in fact paint a picture of the evaluator in this new paradigm:

> When the artist is alive in any person, whatever his kind of work may be, he becomes an inventive, searching, daring, self-expressive creature ….. He disturbs, upsets, enlightens, and opens ways for better understanding. Where those who are not artists are trying to close the book, he opens it and shows there are still more pages possible.
AN ARTFUL APPROACH TO PROGRAM EVALUATION

Endnotes


9. The term stakeholder as used here includes politicians, program management and staff, program clientele, and the public at large.


46. Rossi, Freeman and Wright, 1979, op. cit.
AN ARTFUL APPROACH TO PROGRAM EVALUATION


53. Ibid., p. 220.


56. Renouf, 1979, op. cit., p. 95.


60. Palumbo, 1987, op. cit., p. 35.


63. Ibid.

64. Ibid.


