Smothering Debate, or the Reinvention of Rational State : Evaluation and Evidence

Steve Jacob

2009

PERfEVAL
PERFORMANCE & EVALUATION

www.perfeval.net

THE PERƒEVAL RESEARCH LABORATORY

Laval University's PerƒEval is a research laboratory on the performance and evaluation of public policy. The objective of this pole of expertise and resources is to contribute to discussions on evaluation and to the advancement of knowledge in a field of study in full expansion. Members of this research laboratory conduct theoretical and empirical studies and pay particular attention to both the distribution and use of their results. PerfEval provides training for practitioners and students through teaching and by providing the framework for master's theses and doctoral dissertations. The laboratory offers its specialists and their expertise to public and semi-public organizations with specific training needs or which require counsel in matters of evaluation and performance management. The PerfEval Laboratory team is made up of researchers: professors, professional researchers and graduate students (master's and PhD). These researchers conduct studies around three principal axes: participatory evaluation, performance management and public policy ethics.

THE RESEARCHER

Steve Jacob is an associate professor in the Department of Political Science at Laval University. He is also the Director of the Maîtrise en affaires publiques (MAP) program and President of the Interfaculty Research Ethics Board. Trained as a political analyst and historian, Steve Jacob conducts research on the process of administration modernization and the mechanisms of performance management in Canada and also in several countries of the OECD. Steve Jacob has been an invited professor at several universities: Université libre de Bruxelles, Université catholique de Louvain, Facultés universitaires Saint-Louis, Institut d'études politiques de Toulouse, École doctorale de la francophonie (Bucarest), and Université Senghor de la francophonie (Alexandria). He has also offered several continuing education courses for public servants in Brussels and Tunis. He is a member of several professional and scientific associations. Since 2008, he has been a member of the Administrative Council of the Société québécoise d'évaluation de programmes (SQÉP).

SMOTHERING DEBATE, OR THE REINVENTION OF THE RATIONAL STATE: EVALUATION AND EVIDENCE*

« Opération chloroforme ou la réinvention de l'État rationnel : l'évaluation et les données probantes »,
Criminologie, vol. 42 no.1 (2009) : 201-223

Steve Jacob[1]

Département de science politique [Department of Political Science] Université Laval
Laboratoire de recherche sur la performance et l'évaluation (PerƒEval)
Steve.Jacob@pol.ulaval.ca

Abstract – Policy or program evaluation is rooted in the practices of contemporary public management. In this article, we describe the many aims pursued by evaluation (e.g. information production, program improvement, the fuelling of democratic debate) and the principal debates that drive the evaluation community. Among these debates, the one that focuses on the trend of evidence based policy is one of the more recent and the most lively. The principal tensions focus on the hierarchical organization of evaluative knowledge, the constrained use of knowledge and the will to homogenize evaluative practice. In observing these developments, we also notice a resurgence in the rationalization of decision-making processes and principles of public management that seems to be out of step with the aspirations of the population for further participation and transparency. This article also presents the challenges confronting decision-makers and public administrators when faced with the reinvention of the rational state.

Keywords: Program evaluation, hierarchical organization of knowledge, rationalization of the decision-making process.

INTRODUCTION

Public management in the early 21st century includes trends and reforms that have been trying for years to modernize or even reinvent the state. However, beyond the declarations and appearance of newness, it should not be forgotten that state "reform" is as old as the state itself (Bézès, 2003, 2007). As with other institutions (e.g. the Church, the family, etc.), the state evolves with society. Thus, aspirations of increased transparency and fewer bureaucratic constraints are not solely the result of a "state modernization program," but originate from major themes in contemporary Western societies.

The concern for effectiveness and the appropriate use of resources also affects the workings of the state. This "performance paradigm" (Hamel and Muller, 2007: 131) is accompanied by the promise of better public services that are better defined and better adapted to citizens' needs. In this context, the focus is on intervention quality, problem solving and meeting objectives. According to Carolyn Heinrich (2007), this managerial project is fuelled by two distinct movements that "aim to improve government effectiveness by developing and utilizing a more rigorous base of information and scientific evidence to guide decisions about program design, funding, implementation, and management." (Heinrich, 2007: 256). This movement trend, performance management, emphasizes accountability and monitoring public interventions. Supporters of this approach claim that the knowledge produced can be used, generally speaking, to inform the population or, more particularly, to assess individuals or organizations (e.g. adjusting compensation, as well as awarding bonuses, promotions and contracts). The second movement is evidence-based policy, which initially developed in the health field before spreading to other policy areas. Supporters of this trend encourage the adoption of policies, programs and practices that are supported by reliable data. In this context, the trend dominating contemporary public management can be summed up as, "a *good* policy is a policy that produces the desired effects" (Williams, 2002: 87). Faced with the popularity of this movement, and in the absence of a precise definition of evidence, certain researchers question whether the current enthusiasm is explained by policy development and decision-making having not been previously based on empirical data (Marston and Watts, 2003). The answer to this question is no, and the situation is explained basically by a transformation in the explicit use of knowledge by decision-makers, managers and practitioners (Nutley et al., 2003). In this article, we focus on the return to rationality being presented as an innovation when the strategy actually aims to stifle debate by smothering it under the cover of reform and modernization.

Every generation remodels concepts, governments focus more on the priorities that coincide with their party visions (e.g. re-engineering government, public-private partnerships) and managers invent new working methods that reflect the aspirations of society. Is the evidence-based policy movement the latest incarnation of these currents that are redrawing the boundaries and workings of the state? Initially, the answer appears to be an obvious yes, since the movement undertaken by supporters of this new public management seems to have resulted in radical changes over the past twenty years. Evidence feeds various quality assurance mechanisms (Naughton, 2005). However, when the workings of public administrations are observed more closely and certain public working methods and instruments are examined, this "yes" becomes more relative, as we will see with evaluative practice.

The question we are trying to answer in this article is, "Does evaluation contribute to the evidence-based movement and example-based approaches?" To answer this question, we will concentrate on the ways in which evaluative knowledge is produced and used, as well as on the debates surrounding the development of the evidence-based approach. We will proceed in three steps. First, we will sum up evaluative practices and the diversity that characterizes them. Second, we will focus on the trend that encourages the application of best practices in evaluation and the trend's support among the community of evaluators. Third, we will identify the challenges posed by this return to the rational state.

EVALUATION:
Some Characteristics of a Public Management Tool

A booming practice

Evaluating a program or a public policy entails measuring its effects, and, according to well-defined criteria (e.g. effectiveness, efficiency) making value judgements on these effects, regardless of whether they are desirable or undesirable, direct or indirect, short or long term, etc. The evaluator harnesses methods taken predominantly from the social sciences and uses them to carry out detailed analyses, support his or her assessment of the evaluated subject, and answer questions asked[2] by the sponsor under the terms of reference governing the evaluative process (Jacob, 2004).

Evaluation is traditionally presented as a steering tool for public action whose purpose is to inform the decisions and direct the actions of the state. Evaluative activities have grown significantly in most Western countries over the last 10–15 years (Furubo et al., 2002; Jacob, 2005a), even to the point of being presented as a pressing need by certain governments (Roche, 2005: 308). The Canadian federal government, in seeking to instill a results-based culture, has promoted initiatives in favour of performance measurement and results-based management, thereby increasing the demand for evaluations (TBS, 2005). Evaluation is therefore a fashoinable practice,fueling a thriving market and a growing industry (Leeuw, 2005).

Evaluation is now used in all but a few public policy sectors, although there are differences within sectors in terms of compliance, methodological development, and quality. For example, evaluation of police intervention program results was systematized beginning in the 1990s (Brodeur, 2003; Chemers and Reed, 2005). This development can be explained in part by the need for accountability for management purposes which accompanies the market-based process of modernizing public management (e.g. privatization, contractualization) (Hodge and Coghill, 2007; Leeuw, 2005). Evaluation is sometimes presented as the best method to ensure that money is well-spent (i.e. that public funds are allocated to successful programs), or as the best way to promote organizational learning (Leeuw, 2005). Evaluation seems to naturally foster the managerial paradigm of contemporary government, even if not all evaluations are part of this ideal, since certain authors present evaluation as tool for strengthening social justice (Lincoln, 2003; Mertens, 2003).

A Range of Evaluative Practices and Conceptions

Evaluation has been systematized and structured over the second half of the 20th century through institutional procedures and professional organizations. The history of evaluation is punctuated with passionate debate and characterized by a marked eclecticism, as it is subject to the same epistemological and methodological considerations and reflections that drive the scientific community as whole (Alkin and Christie, 2004).

Today, evaluation brings together a variety of practices, each one very different from the next, which increase the flow of studies and data documenting government operations (Rist and Stame, 2006). The range of perspectives is so wide that it would be impossible to list them all—even more so given that modern societies use the concept of evaluation so extensively that its meaning sometimes risks being watered down. The spectrum ranging from spontaneous evaluation to scientific evaluation is very wide. Successive generations of evaluation come together in an accumulative movement where the advocates of experimental evaluation rub shoulders with the proponents of constructivist evaluation.

Conceptions of evaluation vary from one country to another, and the very notion of evaluation covers a wide range of different realities. Nuances are determined not only by the rationale of a profession (guides, facilitators, critics, judges, etc.), but also by more fundamental attributes. The term "evaluator" describes a professor-researcher, a private consultant and a government agent. This lack of a clear definition of the evaluator's role appears to some to be a weakness which reduces evaluation to an accumulation of subjective judgements or to a practice for lobbying decision-makers (Roche, 2005).

Evaluation also has many purposes that often vary depending on the motivations at the root of the process (Jacob, 2005b). In some cases, an evaluation is conducted because of a legal or regulatory obligation and is therefore a ritual to which decision-makers and managers must subject themselves. In other cases, the evaluation aims to provide information to support the decision-making process or to guide the conduct of public action. Thus, evaluation must reconcile scientific and political dimensions (Taylor, 2005).

---

[2] For example: "Is this policy program reaching its target population? How sound is the underlying 'theory' of this program? Is the intervention being implemented well? How effective is the program? What is the relationship between costs and effects of this program?" (Leeuw, 2005)

Quality questioned

This pluralism in methodological approaches and in the perception of evaluation's purposes, based on different conceptions of or a different importance assigned to the evaluator's role (Alkin and Christie, 2004), regularly sparks debates over methodological rigour and the validity of the knowledge produced. The diverse and heterogeneous nature of evaluative practices appears to some as a source of confusion and threats to evaluation's continuing expansion. Frans Leeuw believes that evaluation's credibility is in danger because "anybody can call themselves an evaluator and bid for evaluation contracts. Sponsors of evaluations and performance reports often lack the expertise to distinguish professional evaluators […] from well-intentioned amateurs or charlatans" (Leeuw, 2005: 238).

Many solutions have been proposed to increase the quality of evaluations (Schwartz and Mayne, 2005). These suggestions range from less restrictive practices, such as adopting a quality charter or standards of good practices (e.g. *Joint Committee on Standards for Educational Evaluation*), to more restrictive measures, such as certifying evaluators, conducting meta-evaluations (evaluations of evaluations) or creating a professional order (Stufflebeam, 2001).

The debate over evidence-based policy in evaluation occurs in the above-mentioned context. More broadly, the popularity of this movement affects evaluation because it generates demands in terms of the practice's sophistication (Martin and Sanderson, 1999), as we will show in the following section.

TOWARDS A HIERARCHIZATION OF EVALUATIVE KNOWLEDGE?

An abundance of data, rather than a lack of it, is currently the primary challenge facing decision-makers, managers and practitioners. With so many channels of information, it is difficult to determine all of the sources available and to assess their rigour and precision. The issue is more one of sorting and hierarchizing knowledge than of scarcity of data (Marston and Watts, 2003).

The resurgence of experimental evaluation

Some authors claim that the issue of quality will only be solved by hierarchizing evaluations based on the Maryland Scientific Methods Scale or on the standards of the Campbell Collaboration,[3] which favour experimental data (Brodeur, 2003; Farrington, 2003; Weisburd and Eck, 2004; Chemers and Reed, 2005; Leeuw, 2005; Greenberg et al., 2006). For these authors, experiments are the best source of evidence, because

> [. . .] randomized experiments provide a higher level of internal validity than do nonexperimental studies. In randomized experiments, people or places are randomly assigned to treatment and control or comparison groups. This means that all causes, except treatment itself, can be assumed to be equally distributed among the groups. Accordingly, if an effect for an intervention is found, the researcher can conclude with confidence that the cause was the intervention itself and not some other confounding factor. (Weisburd and Eck, 2004: 24)

The return to the rational state is accompanied, in evaluation, by a resurgence in experimental methodologies.[4] During the 1960s and 1970s, experimental epistemology structured the development of evaluative practice, including in the security and justice fields (Alkin, 2004; Leeuw, 2005), before losing ground to approaches that broke with positivist views of evaluation. Experimental evaluation associates a restricted number of participants with a process dominated by rationalist concepts (Albaek, 1995) and distinguishes between administrative, scientific and social fields. In reaction to these considerations, new approaches[5] developed that were intended to democratize the production and use of evaluative knowledge to all of the stakeholders concerned by the program being studied in order to put evaluation at the service of the most destitute (Taylor, 2005).

According to proponents of experimentation's resurgence and of works inspired by the Campbell Collaboration, the advantages of the experimental approach are the following: reduced problems with the validity of the data produced; better acceptance of results in the fields of education, administration and politics; accessibility and availability of results and research methods; the presence of conditions favourable to building knowledge and data; and the prospect of transferring knowledge to decision-makers, practitioners, etc. (Leeuw, 2005).

Consequently, experiments are becoming a condition of quality evaluation. Randomized controlled trials (RCTs) are becoming more common in all sectors of public intervention. Very often, the evaluation reports on RCTs are considered to be an obvious, even exclusive, source for systematic reviews. In this regard, the claim by Lawrence W. Sherman that "a review could exclude any evaluation that is not an RCT, as is the common practice in medicine" (2003, 14) demonstrates that experimental approaches increasingly appear as archetypes.[6]

This evaluative approach supports the evidence-based policy movement and tries to position itself as the alternative to the use and quality problems that characterize the history of evaluation. However, while still biased in favour of experimental methods, analyses that compare the conclusions of experimental and non-experimental evaluations show contradictory results. For some analyses, the comparison shows that evaluations of the same program arrive at different conclusions depending on their methodology (Sherman, 2003), while other analyses conclude that the results are relatively the same (Greenberg et al., 2006).

---

[3] Founded in 2000, the Campbell Collaboration is an international organization that aims to study the effectiveness of social and behavioural interventions. See http://www.campbellcollaboration.org/

[4] This process has been used in Quebec in particular to assess the effects of a break-and-enter prevention program (Charest, 2003) and an intensive probation program at the Centre jeunesse de Montréal [Montreal Youth Centre] (Cournoyer and Dionne, 2007).

[5] These approaches do not have a blanket term. The concept of democratic evaluation includes participative evaluations, fourth-generation evaluations, emancipative evaluations, etc.

[6] In this regard, the American Evaluation Association has published a critical opinion concerning the desire to consider RCTs as the best or only reliable method.
See http://www.eval.org/doestatement.htm (consulted on February 10, 2009).

Restricted use of knowledge

Behind the notion of evidence-based policy lies a sort of automatism in the decision-making process and the conduct of public action. The process can be summed up in a statement such as "The data tell us that . . . and we must act accordingly." This trend seeks to bring together the knowledge producers (e.g. the evaluators, researchers and academics) with the decision-making centres (Marson and Watts, 2003: 147). In this context, decision-makers, and sometimes even voters, are seen as clients whom the evaluator or researcher must satisfy (Shulock, 1999; Sherman, 2003). According to a "rationalist model, there should be a clear separation between knowledge production (by experts) and knowledge use (by practitioners)" (Taylor, 2005: 606). It is not enough to speak to decision-makers; they must also listen. In evaluation, the issue of how results are used is a central concern. An evaluation's conclusions and recommendations do not systematically lead to changes on the ground. Therefore, with a few rare exceptions, evaluation has only limited influence on decision-makers and public managers (Weiss et al., 2008). While the managerial paradigm increases the number of reports to be written (annual, activity and evaluation), most of these documents seem to be read more before they are published than after (Jacob, 2006).

Research into how evaluation results are used forces us to relativize the impact that evaluators and researchers (in social sciences) can have on public decision-makers (Shulock, 1999). The issues of the relevance and validity of data to be used are at the heart of this consideration. This issue has two opposing viewpoints. Some consider than a considerable amount of knowledge has already been produced, but that it is poorly known or unknown by decision-makers. The preferred way to improve the situation is to democratize or popularize this knowledge without trying to hierarchize it, and to assist potential users to easily apply this knowledge to their specific situations (Weiss et al., 2008). The other side of the debate claims that efforts must be made to promote evidence-based policy. For this side, the problem of not using knowledge can be overcome by improving the quality of the knowledge produced and by training of decision-makers to be "intelligent consumers of complex evidence" (Sanderson, 2002: 6). To accomplish this goal, "reliable," "valid" and "solid" data must be given precedence (Heinrich, 2007). The latter movement has been increasingly popular in the past few years, including among decision-makers and within the evaluator community. For example, in the United States, the 1993 *Governance Performance Results Act* (GPRA) and the more recent *Program Assessment Rating Tool* (PART) promote the production and use of evidence to evaluate federal program effectiveness (Chemers and Reed, 2005). These evaluations are also used in budgetary discussions, because the Office of Management and Budget (OMB) gives greater priority to programs that document their effectiveness using experimental evaluations (Heinrich, 2007). In short, "GPRA and PART have fuelled a need to understand evaluation concepts and, more importantly, to think in evaluative terms" (Chemers and Reed, 2005: 263).

The movement taking shape is characterized by restricted use of evaluative knowledge. A program's funding is contingent upon showing its effectiveness using rigorously scientific evaluation (Weiss et al., 2008). This restricted use has been present for many years in the international aid sector, where donors require accountability reporting by recipient countries. The current trend is towards generalizing this practice to the fields of prevention (e.g. drug abuse, at-risk sexual behaviour, violence, etc.) or to all public programs. This situation raises new challenges for managers. In a context where funding is conditional upon using scientifically valid data, how can local agencies or services with limited staff produce this knowledge themselves? Based on the evaluation of a drug abuse prevention program in American schools, Carol Weiss et al. (2008) find that most organizations, including the largest, turn to lists of scientifically proven programs. This "victory for evaluation [. . .] rests on the assumptions that good evidence was collected and assembled, and that the list of proven and promising programs was a fair and accurate representation of the evidence" (Weiss et al.: 38). In their analyses, these researchers question the above assumptions and show that the evaluations of programs on these lists are not always without methodological shortcomings (e.g. the groups compared deal with different versions of the program, the studies were carried out soon after the program's implementation) or ethical shortcomings (e.g. possible favouritism or conflict of interest when the evaluation is carried out by the program designer or by an expert who designed an alternative program). The researchers conclude that the evidence-based requirement is more of a new bureaucratic hurdle than a step forward in the use of evaluative knowledge. Increasingly often, thinking about how to produce the knowledge required to make decisions and analyzing the data provided is being supplanted by mechanically using the results of experiments and systematic reviews of the literature. This situation poses new challenges.

For example, the increase in accountability and performance management requirements also causes negative effects on administrative services. Several studies explain that managers and stakeholders adapt their work according to performance criteria. More attention is sometimes paid to how to "appear to be performing" than to how to satisfy user needs. The practices of falsifying measurement tools and manipulating data spread as performance monitoring tools develop (van Thiel and Leeuw, 2002; Heinrich, 2007).

THE CHALLENGES OF PUBLIC RATIONALIZATION

As we have stated many times in this article, the trend emphasizing the use of evidence is a "new" way of rationalizing public management that has its limits. In this last section, we will review a few general criticisms of the evidence-based movement from the literature. Space constraints prevent us from providing an exhaustive list. To open the debate, we will highlight the general issues raised by the return to the rational state.

Managing power issues

Evidence-based public management is a "new" method of political regulation (Taylor, 2005) to which evaluation contributes. This movement is developing in an increasingly ideology-free environment, where evidence takes precedence over political positions. This tendency to depoliticize evidence is concomitant with other developments in public management, such as the "displacement of responsibility from public decision making to independent public agencies" (Walker, 2007: 236). However, developing and implementing policies and programs remains a political process, during which values enter into consideration (Marston and Watts, 2003; Naughton, 2005). It is unrealistic to think that the data produced is neutral (Williams, 2002) or that it will not be used or misused as ammunition in the public debate, since politics is characterized by confrontational arguments (Albaek, 1995; Hanberger, 2006). However, "a lack of empirical evidence supporting opinions does not mean that all the opinions are wrong" (Chalmers, 2003: 24). It seems far-fetched to think that evidence alone allows the decision-making process to be depoliticized.

Scientific and technocratic trends tend to obscure the fact that evaluation is not always a neutral method of producing knowledge. Evaluation involves power issues that must not be forgotten. Evaluation can have an impact on public decision making, can have repercussions on how the administrative management of programs is handled, and can even trigger a reconsideration or reinterpretation of the problems that lead to public intervention. Evaluation is also a strategic tool to organize or even steer public debate on current issues. In this context, the evaluator must confront the inertial forces that try to maintain the status quo and the forces of change that favour a program of reforms. These forces are fuelled by partisan, ideological motivations, but also by which methods of intervention are preferred by administrations and professional associations, and by the demands of the public, interest groups and user associations (Davies et al., 1999; Weiss et al., 2008). To sum up, the current debate on the production and use of evidence updates older arguments about the diversity and validity of knowledge through the "devaluation of other forms and sources of knowledge, including the input of ordinary citizens" (Heinrich, 2007: 269).

Overcoming methodological difficulties

Although the scientific ambitions claimed by promoters of the evidence-based policy movement are clear, applying these ambitions in a concrete setting raises several methodological difficulties.[7]

Many different expectations and differing understandings of the issues very often appear when one goes into the field and speaks with several actors involved in developing and implementing a program (elected officials, managers, practitioners, etc.). Certainties break down as complexity increases. For example, various services and practitioners sometimes have difficulty identifying the relevant unit of analysis, or even precisely defining the outline of a program. In this context, it becomes difficult to be able to understand all of the phenomena and to account for them in detail. To produce evidence, the situation must often be simplified.

Many researchers and evaluators agree on the difficulty inherent in wanting to understand the entirety of a situation. This task is exceedingly complex and often requires a combination of several disciplines (Jacob, 2008) or an analytical simplification. As we have seen, RCTs focus primarily on program results in order to identify successful programs by answering the question of "what works?" Despite its growing popularity in the United States, this approach is not without criticism, as we can see by examining the following three issues.

---

[7] In this section, we will limit ourselves to the methodological difficulties present in Western countries. However, the evidence-based policy movement hopes to spread its working methods throughout the world. Readers interested in the issues faced by developing countries can consult the study by Bowles et al., 2005.

The approach is criticized for its narrow field of view, that is limiting its observations and analyses to program effects or to elements that can be measured (Williams, 2002). However, any evaluation must deal with the difficult necessity of controlling the complexity of the social world and of the environment in which the program is implemented (the "why?"). This need to put public intervention into context is necessary to adequately interpret the results produced and to understand "why programs (do not) work." For example, a lack of expected results is not always synonymous with program failure and could be explained by other reasons, such as a shortcoming in how the program was implemented. Some believe that evaluation cannot limit itself to only the results, but that it is vital to understand the process that influences the execution of a program in order to better judge the program's relevance. In some cases, adjustments in a program's implementation are sufficient to generate the desired effects (Oakley et al., 2004). In the same vein, it has been shown that negative evaluation results are fairly often disguised. Less flattering conclusions for an organization or manager are rarely presented at large international conferences and are not often published in scientific journals. They are less likely to be disseminated and then cited in other studies, including in systematic literature reviews (Chalmers, 2003: 25).

Next, more and more evaluations focus primarily on the target group, namely the actor considered to be the source of the problem (e.g. street gangs, impaired drivers, "inconsiderate" residents in a residential neighbourhood). This choice (the "who?") is criticized by several authors, who believe that using a single actor as a benchmark for public interventions is simplistic, given that program effects stem from interactions between a variety of actors whose behaviours affect the overall results (Williams, 2002).

Other observers believe that the subject studied (the "what?") is too narrow to be able to understand the changes in behaviour caused by public programs (Rossi et al., 2004; Chen, 2005). Public intervention is based on a theory of social change. Taking this theory into consideration is vital for proponents of realistic evaluation (Pawson, 2002; Pawson and Tilley, 1994). In their opinion, we must look outside the administrative black box to fully understand the process at work during program implementation. However, this theory can be poorly formulated, lead to substitution effects, or even do more harm than good (Chalmers, 2003). By having too limited and too narrow a vision of public action, it is difficult for supporters of the evidence-based approach to take into consideration the inevitable unforeseen effects that characterize state intervention. Moreover, policy analysts know that the process of implementing a policy does not always take place as the decision-makers expected (Pressman and Wildavsky, 1984) and that "adequately" implemented public interventions sometimes produce unexpected effects (desirable and undesirable). For example, fighting graffiti on subway trains by increasing (video) surveillance and using a cleaning crew may meet its objective but displace the problem by forcing taggers to choose other, less well-monitored areas. This is why managers of different sectors of public policy have difficulty arriving at a shared vision of success. Success for some is sometimes the beginning of the problem for others. Interventions working in isolation, silo-style management and the lack of attention to crafting coherent public policies and programs weaken the pursuit and the improvement of the general public interest. Operational and realistic motivations justify deconstructing complex issues into more focussed problems for the purposes of intervention. However, at some point, actions must be reconsidered from a wider perspective, and competing objectives must be arbitrated between. This task is essentially political and can draw on knowledge produced by evaluations without being bound to follow evaluation recommendations because "in the eyes of ministers, ideology, opportunity and public preferences compete with scientific knowledge" (Davies et al., 1999: 4). Based on a Dutch crime prevention example, Frans Leeuw et al. (2007) believe that, by combining evaluation approaches applicable to the theory of the program with the methodological requirements promoted by the Campbell Collaboration, the conclusions become more credible and accessible, as well as more likely to be used by decision-makers.

Finally, one of the challenges any evaluation must deal with involves defining the right time (the "when?") to assess all of the effects (intentional, unintentional, desirable and undesirable) produced by public intervention. By starting an evaluation too early, the evaluator risks being unable to observe the changes a program is expected to cause and therefore concluding too soon that the program is a failure. Inversely, by starting the evaluation too late, the evaluator risks no longer being able to identify the program's effects. Timing when to start the evaluation is crucial and affects the results. This situation applies in fighting crime, where "various prevention measures can take varying amounts of time to affect crime" (Charest, 2003: 34 [TR note: unofficial translation]). There is no hard-and-fast rule to determine precisely when all of the effects of the intervention will become apparent. The evaluator must reconcile various time-sensitive requirements (annual budget cycle, duration of the legislature, multi-year planning, etc.) and be aware of the limits of the analysis.

The performance management paradigm increases short-term monitoring requirements and reconfigures how services operate. Managers encourage actions that result in quickly observable changes, to the detriment of other initiatives that produce change more slowly. Furthermore, some decision-makers are in situations where it is difficult to anticipate new challenges and emerging pressures that require interventions which are speedy, innovative and adapted to specific environments. Aside from the traditional methodology issues, the requirements formulated by proponents of the evidence-based movement fall within this context without really questioning the validity of the results produced. However, as Frank Furstenberg (2003), a researcher who has studied single mothers for 35 years, explains, conclusions vary according to the periods analyzed. He believes that the conclusions he reached after a 12-year follow-up would have been very different if he had ended his analysis after five years.

Being compatible with the other "requirements" of contemporary public management

Concern for performance accompanies contemporary public management modernization processes. Decision-makers and managers must not only make the right choices, they must show citizens (often through the media) that they have made the right decisions and that their actions resulted in a substantial improvement to the situation. In this context, the evaluator naturally appears to be an "expert" (Hanberger, 2006) or an "evidence bearer" (Taylor, 2005: 605).

Aside from the appearance of newness created by this trend, it is interesting to note that evidence-based policy is contributing to the reinvention of the rational state (Sanderson, 2004). A few decades ago, making rational decisions consisted of identifying a problem, listing all of the possible options for solving it and finding the best cost-benefit ratio (Williams, 2002). It should be recalled that this decision-making method had been questioned, and technocrats and elitist governance had been denounced. The following movement questioned public expertise and top-down decisions that did not take into account the specifics of where measures were implemented. Evaluation, and more broadly, scientific research, is having this debate again because, on one side, those who focus on producing and using evidence believe that evaluation should be addressed primarily to decision-makers and managers (Pawson, 2002), while others, such as Anders Hanberger (2001), think that evaluation must directly serve the people. According to the latter researcher, evaluative knowledge must above all serve to hold institutions accountable and to provide information to improve democratic debate and decision making. Evaluation should not be addressed only to a monolithic public, but bridges should be built between the various groups that make up society (Hanberger, 2001: 212). From this perspective, evaluation is a mediation and conciliation tool that breaks with the managerial and technocratic visions that drive the evidence-based movement, and coexists with an increased demand for participation by citizens and non-governmental actors in the development and implementation of an increasing number of public programs (OECD, 2001; Williams, 2002). Evaluation is no longer a top-down movement, but also a tool that citizens can use to convey information back up. Therefore, the participatory dimension of evaluation expands the "expert" vision of knowledge production to new actors who are not all seen as legitimate by certain decision-makers and evaluators.

Furthermore, public managers and decision-makers must take into account a growing ethical requirement. Moreover, experiments often raise ethical questions in their implementation. In some cases, depending on the nature of the interventions upon which the experiment is based, a feeling of favouritism can occur in a portion of the population (Bowles et al., 2005). In other cases, certain experiments are more hazardous and can have harmful consequences on the citizens subjected to them. While research ethics are developing to limit these risks and to weigh the resulting advantages compared with the inconvenience suffered by participants, experiments are not systematically subject to standards and principles of ethical research.[8]

Finally, the reforms promoted by the theorists of the new style of public management involve changing the citizen-user into the citizen-client. The latter becomes the centre of state functioning and is subject to increased attention. This focus explains the development of user-satisfaction studies, which also produce knowledge that must be integrated into decision making. Moreover, public expectations can be satisfied without necessarily meaning that the interventions are effective in the sense agreed upon by proponents of the evidence-based movement. Furthermore, the adoption of public programs and policies must be supported by a principle of realism in terms of how planned solutions would be received and whether they could be implemented (Naughton, 2005). In this context, what data should take precedence in decision making? Who will make the necessary decisions and according to what principles? These are the questions that must be answered to reconcile the sometimes contradictory requirements at the heart of contemporary public management.

---

[8] For example, in Canada, the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans states that quality assurance procedures and evaluations should not be studied by a research ethics committee (Inter-Agency Secretariat on Research Ethics, 2005).

CONCLUSION

As we have seen, and to answer the question we asked in the introduction, certain forms of evaluation fuel the evidence-based movement. The community of evaluators is so varied that it includes both supporters and detractors of this approach. The differences can be explained by evaluators' training, the sectors in which they work and the sponsor requirements regarding the type of evaluations they perform. There is no consensus on the subject and, as shown by the history of evaluation, consensus has regularly adapted to changes in public management and has often developed based on arguments about modernizing the state.

Aside from the theoretical reflections that we have presented in this article, it must be remembered that evaluation is also a useful practice for practitioners working in very diverse fields (e.g. crime prevention, correctional services, youth protection, public safety, etc.). The knowledge produced by evaluation allows practitioners to understand their actions differently and to more accurately understand the ins and outs of their situations. In short, evaluation provides a break from the "urgency of the everyday" that contributes to critical thinking, and individual and organizational learning that allows interventions to be put into perspective and examined in their entirety. From then on, each field can ask itself what type of knowledge is the most appropriate to fuel its reflection, based on the diversity of evaluative practices that we have presented in this article to inform them.

In conclusion, evidence-based approaches have the virtue of re-energizing the debates about methodologies used by evaluators and about how decision-makers, managers and practitioners use the evaluator's conclusions and recommendations. By taking into consideration the challenges of rationalizing public action that we have just presented, the contributions of supporters and detractors of these practices can be enriched and reconsidered according to the benchmark of the needs and constraints of the societies in which they are developing. During this reflection, it must be remembered that researchers have always had difficulty finding their place in the relationship between political will and administrative management (Walker, 2007). As Keynes once observed, "there is nothing a government hates more than to be well-informed; for it makes the process of arriving at decisions much more complicated and difficult" (Skidelsky, 1980: 630).

REFERENCES

Alback, E. (1995). Between knowledge and power: utilization of social science in public policy making.
Policy Sciences, 28, 79-100.

Alkin, M. C. (Ed.) (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks: Sage.

Alkin, M. C., & Christie, C. A. (2004). An Evaluation Theory Tree. In M. C. Alkin (ed.),
Evaluation Roots: Tracing Theorists' Views and Influences (12-65). Thousand Oaks: Sage.

Bézès, P. (2003). Aux origines des politiques de réforme administrative sous la Vème République :
la construction du « souci de soi de l'État ». *Revue française d'administration publique* 102, 307-325.

Bézès, P. (2007). Construire des bureaucracies wébériennes à l'ère du
New Public Management ? Critique internationale, 35, 9-29.

Bowles, R., Akpokodje, J., & Tigere, E. (2005). Evidence-based approaches to crime prevention in developing countries.
European Journal on Criminal Policy and Research, 11, 347-377.

Brodeur, J.-P. (2003). À la recherche d'une évaluation « pauvre ». *Criminologie, 36* (1), 9-30.

Chalmers, I. (2003). Trying to do more good than harm in policy and practice: the role of rigorous, transparent,
up-to-date evaluations. *Annals of the American Academy of Political and Social Science*, 589, 22-40.

Charest, M. (2003). Effets préventifs et dissuasifs : analyse d'impact d'une opération policière
de prévention des cambriolages. *Criminologie, 36* (1), 31-56.

Chemers, B. & Reed, W. (2005). Increasing evidence-based programs in criminal and juvenile justice:
a report from the front lines. *European Journal on Criminal Policy and Research*, 11, 259-274.

Chen, H.-T. (2005). *Practical program evaluation*. Thousand Oaks: Sage.

Cournoyer, L.-G., & Dionne, J. (2007). Efficacité du programme de probation intensive
du Centre jeunesse de Montréal-Institut universitaire : la récidive officielle. *Criminologie, 40* (1), 155-183.

Davies, H. T. O., Nutley, S. M., & Smith, P. C. (1999). Editorial: What Works?
The Role of Evidence in Public Sector Policy and Practice. *Public Money and Management, 19* (1), 3-5.

Farrington, D. P. (2003). Methodological quality standards for evaluation.
The Annals of the American Academy of Political and Social Science, 587, 49-68.

Furstenberg, F. (2003). Teenage childbearing as a public issue and private concern. *Annual Review of Sociology*, 29, 23-39.

Furubo, J.-E., Rist, R., & Sandahl, R. (Eds.), (2002). *International Atlas of Evaluation*.
New Brunswick and London: Transaction Publishers.

Greenberg, D. H., Michalopoulos, C., & Robins, P. K. (2006). Do experimental and nonexperimental
evaluations give different answers about the effectiveness of government-funded training programs?
Journal of Policy Analysis and Management, 25 (3), 523-552.

Hamel, M.-P., & Muller, P. (2007). L'accès aux droits sociaux : un compromis entre performance gestionnaire
et justice sociale. *Politiques et management public, 25* (3), 131-149.

Hanberger, A. (2001). Policy and Program Evaluation, Civil Society, and Democracy.
American Journal of Evaluation, 22 (2), 211-228.

▍**Hanberger, A.** (2006). Evaluation of and for democracy. *Evaluation, 12* (1), 17-37.

▍**Heinrich, C. J.** (2007). Evidence-Based Policy and Performance Management.
Challenges and Prospects in Two Parallel Movements. *The American Review of Public Administration, 37* (2), 255-277.

▍**Hodge, G. A.,** & Coghill, K. (2007). Accountability in the Privatized State.
Governance: An International Journal of Policy, Administration and Institutions, 20 (4), 675-702.

▍**Inter-Agency Secretariat on Research Ethics** (2005). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.* Ottawa: Public Works and Government Services Canada.

▍**Jacob, S.** (2004). Évaluation. In L. Boussaguet, S. Jacquot & P. Ravinet (eds.),
Dictionnaire des politiques publiques (201-208). Paris : Presses de la Fondation Nationale des Sciences Politiques.

▍**Jacob, S.** (2005a). *Institutionnaliser l'évaluation des politiques publiques. Étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas.* Bruxelles : PIE-Peter Lang.

▍**Jacob, S.** (2005b). Réflexions autour d'une typologie des dispositifs institutionnels d'évaluation.
Revue canadienne d'évaluation de programme, 20 (2), 49-68.

▍**Jacob, S.** (2006). Trente ans d'évaluation de programme au Canada : l'institutionnalisation interne en quête de qualité.
Revue française d'administration publique, 119, 515-532.

▍**Jacob, S.** (2008). Cross-Disciplinarization: A New Talisman for Evaluation? *American Journal of Evaluation, 29* (2), 175-194.

▍**Leeuw, F. L.** (2005). Trends and Developments in Program Evaluation in General and Criminal Justice Programs in particular.
European Journal on Criminal Policy and Research, 11, 233-258.

▍**Leeuw, F. L.,** van der Knaap, L. M., & Bogaerts, S. (2007). Reducing the knowledge-practice gap: a new method
applied to crime prevention. *Public Money and Management, 27* (4), 245-250.

▍**Lincoln, Y. S.** (2003). Fourth Generation Evaluation in the New Millennium. In S. I. Donaldson & M. Scriven (eds.),
Evaluating Social Programs and Problems: Visions for the New Millenium (77-90). Mahwah, NJ: Erlbaum.

▍**Marston, G., & Watts, R.** (2003). Tampering with the evidence: a critical appraisal of evidence-based policy making.
The Drawing Board: An Australasian Review of Public Affairs, 3 (3), 143-163.

▍**Martin, S., & Sanderson, I.** (1999). Evaluating public policy experiments. Measuring outcomes,
monitoring processes or managing pilots? *Evaluation, 5* (3), 245-258.

▍**Mertens, D. M.** (2003). The Inclusive View of Evaluation: Visions for the New Millenium. In S. I. Donaldson & M. Scriven (eds.),
Evaluating Social Programs and Problems: Visions for the New Millenium (91-108). Mahwah, NJ: Erlbaum.

▍**Naughton, M.** (2005). "Evidence-based policy" and the government of the criminal justice – only if the evidence fits!
Critical Social Policy, 25 (1), 47-69.

▍**Nutley, S. M., Walter, I., & Davies, H. T. O.** (2003). From knowing to doing.
A framework for understanding the evidence-into-practice agenda. *Evaluation, 9* (2), 125-148.

▍**Oakley, A., Strange, V., Stephenson, J., Forrest, S., & Monteiro, H.** (2004). Evaluating Processes.
A Case Study of a Randomized Controlled Trial of Sex Education. *Evaluation, 10* (4), 440-462.

▍**OECD** (2001). *Des citoyens partenaires : information, consultation, et participation à la formulation des politiques publiques*. Paris : OECD.

▍**Pawson, R.** (2002). Evidence and Policy and Naming and Shaming. *Policy Studies, 23* (3-4), 211-230.

▍**Pawson, R., & Tilley, N.** (1994). What works in evaluation research? *British Journal of Criminology, 34* (3), 291306.

▍**Pressman, J. L., & Wildavsky, A.** (1984). *Implementation*. Berkeley: University of California Press.

▍**Rossi, P. H., Lipsey, M. W., & Freeman, H. E.** (2004). *Evaluation: A Systematic Approach* (7th ed.). Thousand Oaks: Sage.

▍**Rist, R., & Stame, N.** (Eds.) (2006). *From Studies to Streams. Managing Evaluative Systems*. New Brunswick & London: Transaction Publishers.

▍**Roche, S.** (2005). The transferability of evaluation and the "what works" approach in France. *European Journal on Criminal Policy and Research*, 11, 297-320.

▍**Sanderson, I.** (2002). Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration, 80* (1), 1-22.

▍**Sanderson, I.** (2004). Getting evidence into practice. Perspective on rationality. *Evaluation, 10* (3), 366-379.

▍**Schwartz, R., & Mayne, J.** (Eds.) (2005). *Quality Matters: Seeking Confidence in Evaluating, Auditing, and Performance Reporting*. New Brunswick, NJ: Transaction Publishers.

▍**Sherman, L. W.** (2003). Preface: Misleading evidence and evidence-led policy: making social science more experimental. *Annals of the American Academy of Political and Social Science*, 589, 6-19.

▍**Shulock, N.** (1999). The paradox of policy analysis: if it is not used, why do we produce so much of it? *Journal of Policy Analysis and Management, 18* (2) 226-244.

▍**Skidelsky, R.** (1980). *John Maynard Keynes* (Vol. 2). London: Macmillan.

▍**Stufflebeam, D. L.** (2001). The Metaevaluation Imperative. *American Journal of Evaluation, 22* (2), 183-209.

▍**Taylor, D.** (2005). Governing Through Evidence: Participation and Power in Policy Evaluation. *Journal of Social Policy, 34* (4), 601-618.

▍**van Thiel, S., & Leeuw, F. L.** (2002). The performance paradox in the public sector. *Public Performance and Management Review, 25* (3), 267-281.

▍**Treasury Board of Canada Secretariat** (2005). *The Health of the Evaluation Function in the Government of Canada. Report for Fiscal Year 2004-05*. Ottawa: TBS.

▍**Walker, D.** (2007). Is Evidence for Policy Good for Democracy? *Public Money and Management, 27* (4), 235-237.

▍**Weisburd, D., & Eck, J. E.** (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science*, 593, 42-65.

▍**Weiss, C. H., Murphy-Graham, E., Petrosino, A., & Gandhi, A. G.** (2008). The Fairy Godmother – and Her Warts. Making the Dream of Evidence-Based Policy Come True. *American Journal of Evaluation, 29* (1), 29-47.

▍**Williams, N.** (2002). Evidence and Policy: Towards a New Politico-Administrative Settlement. *The Political Quarterly, 73* (1), 86-97.