

Opération chloroforme ou la réinvention de l'État rationnel : l'évaluation et les données probantes

Steve Jacob

Volume 42, numéro 1, printemps-été 2009

L'intervention à l'aune des données « probantes »

URI : <https://id.erudit.org/iderudit/029813ar>

DOI : <https://doi.org/10.7202/029813ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0316-0041 (imprimé)

1492-1367 (numérique)

[Découvrir la revue](#)

Citer cet article

Jacob, S. (2009). Opération chloroforme ou la réinvention de l'État rationnel : l'évaluation et les données probantes. *Criminologie*, 42(1), 201–223.
<https://doi.org/10.7202/029813ar>

Résumé de l'article

L'évaluation de politiques ou de programmes s'est enracinée dans les pratiques de la gestion publique contemporaine. Dans le présent article, nous décrivons les multiples finalités poursuivies par l'évaluation (ex. : production d'information, amélioration des programmes, alimentation du débat démocratique) et les principaux débats qui animent la communauté des évaluateurs. Parmi ces débats, celui sur le courant de la décision fondée sur des données probantes est l'un des plus récents et des plus vifs. Les principales tensions portent sur la hiérarchisation des connaissances évaluatives, l'utilisation contrainte des connaissances et la volonté d'homogénéiser la pratique évaluative. En observant ces développements, nous constatons également une résurgence de la rationalisation des processus de prise de décisions et des principes de gestion publique qui semblent être en décalage avec les aspirations de la population à plus de participation et de transparence. Dans le présent article, nous présentons également les défis auxquels sont confrontés les décideurs et les gestionnaires publiques face à la réinvention de l'État rationnel.

Opération chloroforme ou la réinvention de l'État rationnel : l'évaluation et les données probantes

Steve Jacob¹

Département de science politique, Université Laval

Laboratoire de recherche sur la performance et l'évaluation (PerfEval)

Steve.Jacob@pol.ulaval.ca

RÉSUMÉ • *L'évaluation de politiques ou de programmes s'est enracinée dans les pratiques de la gestion publique contemporaine. Dans le présent article, nous décrivons les multiples finalités poursuivies par l'évaluation (ex. : production d'information, amélioration des programmes, alimentation du débat démocratique) et les principaux débats qui animent la communauté des évaluateurs. Parmi ces débats, celui sur le courant de la décision fondée sur des données probantes est l'un des plus récents et des plus vifs. Les principales tensions portent sur la hiérarchisation des connaissances évaluatives, l'utilisation contrainte des connaissances et la volonté d'homogénéiser la pratique évaluative. En observant ces développements, nous constatons également une résurgence de la rationalisation des processus de prise de décisions et des principes de gestion publique qui semblent être en décalage avec les aspirations de la population à plus de participation et de transparence. Dans le présent article, nous présentons également les défis auxquels sont confrontés les décideurs et les gestionnaires publiques face à la réinvention de l'État rationnel.*

MOTS-CLÉS • *Évaluation de programmes, hiérarchisation des connaissances, rationalisation de la prise de décisions.*

Introduction

La gestion publique de ce début de ^{xxi}e siècle témoigne des évolutions et des réformes qui, depuis de nombreuses années, ambitionnent de

1. Je remercie Denis Lafortune de m'avoir offert l'occasion de participer à ce numéro thématique et les évaluateurs de l'article pour leurs commentaires et suggestions. Cet article a été commenté par des membres du Laboratoire de recherche sur la performance et l'évaluation (PerfEval) de l'Université Laval; qu'ils soient ici remerciés.

moderniser, voire de réinventer l'État. Cependant, au-delà des déclarations et de l'apparence de nouveauté, il ne faut pas perdre de vue que la « réforme » de l'État est aussi ancienne que l'État lui-même (Bézès, 2003, 2007). L'État, au même titre que d'autres institutions (ex. : l'Église, la famille, etc.), évolue au rythme des sociétés. Ainsi, les aspirations à plus de transparence et à moins de contraintes bureaucratiques ne s'expliquent pas seulement par l'application d'un « programme de modernisation de l'État », mais trouvent leurs origines dans de grands courants qui traversent les sociétés occidentales contemporaines.

Le souci d'efficacité et d'utilisation appropriée des ressources affecte également le fonctionnement de l'État. Ce « paradigme de la performance » (Hamel et Muller, 2007 : 131) s'accompagne de la promesse de meilleurs services publics, mieux définis et plus adaptés aux besoins des citoyens. Dans ce contexte, l'accent est mis sur la qualité des interventions, la résolution des problèmes ou l'atteinte des objectifs. Selon Carolyn Heinrich (2007), ce projet managérial est alimenté par deux courants distincts qui ambitionnent « l'amélioration de l'efficacité de l'État grâce à la production et à l'utilisation d'informations et de connaissances scientifiques rigoureuses pour orienter les décisions relatives à la conception, au financement, à la mise en œuvre et à la gestion des programmes » (Heinrich, 2007 : 256, trad.). Il s'agit tout d'abord du courant de la gestion centrée sur la performance (*performance management*) qui met l'accent sur le suivi des interventions publiques et la reddition de comptes. Pour les partisans de cette approche, les connaissances produites peuvent être utilisées, d'un point de vue général, pour informer la population ou, plus particulièrement, pour apprécier les individus ou les organisations (ex. : adaptation de la rémunération, attribution de primes, de promotions ou de contrats). Le deuxième courant est celui de la décision fondée sur des données probantes (DP) (*Evidence Based Policy*) qui s'est initialement développé dans le domaine de la santé avant de s'étendre aux autres secteurs de politique. Les promoteurs de ce courant favorisent l'adoption de politiques, de programmes ou de pratiques qui sont soutenus par des données fiables. Dans ce contexte, la tendance qui domine la gestion publique contemporaine peut se résumer par l'affirmation suivante : « une *bonne* politique est une politique qui produit les effets désirés » (Williams, 2002 : 87, trad.). Devant la popularité de ce courant, et en l'absence d'une définition précise du concept de DP, certains chercheurs se demandent si l'enthousiasme actuel se justifie par le fait que l'élaboration de politiques

et la prise de décisions n'étaient pas auparavant basées sur des données empiriques (Marston et Watts, 2003). La réponse à cette question est négative et il s'agit en fait essentiellement d'une transformation dans l'utilisation explicite des connaissances par les décideurs, les gestionnaires et les intervenants (Nutley *et al.*, 2003). Dans le présent article, nous mettons l'accent sur le retour de la rationalité, présenté comme une innovation, qui s'apparente à une opération chloroforme, c'est-à-dire à une entreprise qui vise à apaiser les débats sous le couvert de réforme et de modernisation.

Chaque génération remodèle les concepts, les gouvernements mettent davantage l'accent sur les priorités qui coïncident avec leurs visions partisans (ex. : la réingénierie de l'État, les partenariats public-privé) et les gestionnaires inventent de nouvelles façons de faire qui reflètent les aspirations de la société. Le courant en faveur de la décision fondée sur des DP est-il le dernier avatar de ces tendances qui redessinent les frontières et le fonctionnement de l'État ? À première vue, la réponse apparaît évidente et positive, tant le mouvement entrepris par les tenants de la nouvelle gestion publique au cours des vingt dernières années semble induire des changements radicaux. Les DP alimentent différents mécanismes d'assurance qualité (Naughton, 2005). Toutefois, lorsque l'on observe le fonctionnement des administrations publiques de plus près et que l'on s'intéresse à certains instruments ou modes d'action publique, il est nécessaire de relativiser ce jugement, comme nous allons le voir avec la pratique évaluative.

La question à laquelle nous tentons de répondre dans le présent article est : « L'évaluation contribue-t-elle au mouvement centré sur les DP et les approches exemplaires ? » Pour y parvenir, nous nous concentrons sur les mécanismes de production et d'utilisation de connaissances évaluatives, ainsi que sur les débats qui entourent le développement de l'approche centrée sur les DP. Nous procédons en trois étapes. Tout d'abord, nous brossons un portrait de la pratique évaluative et de la diversité qui la caractérise. Ensuite, nous nous intéressons au courant qui promeut l'application de pratiques exemplaires en évaluation et de son adhésion au sein de la communauté des évaluateurs. Enfin, nous identifions les défis que pose ce retour de l'État rationnel.

L'évaluation : quelques caractéristiques d'un instrument de gestion publique

Une pratique en plein essor

Évaluer un programme ou une politique publique consiste à mesurer les effets propres à cette politique et, en fonction de critères bien définis (ex. : efficacité, efficience, etc.), à porter un jugement de valeur sur ces effets, qu'ils soient voulus ou pervers, directs ou indirects, induits à court ou à long terme. L'évaluateur mobilise des méthodes provenant, notamment, des sciences sociales pour conduire des analyses détaillées, étayer son appréciation de l'objet évalué et répondre aux questions posées² par le commanditaire dans le cahier des charges à l'origine de la démarche évaluative (Jacob, 2004).

L'évaluation est traditionnellement présentée comme un instrument de pilotage de l'action publique qui vise à éclairer la décision et à orienter les interventions de l'État. En comparaison avec la situation des dix ou quinze dernières années, les activités évaluatives se sont considérablement développées dans la plupart des pays occidentaux (Furubo *et al.*, 2002 ; Jacob, 2005a). Elles sont même présentées par certains gouvernements comme une « ardente obligation » (Roche, 2005 : 308). Au Canada, le gouvernement fédéral, souhaitant s'inscrire dans une culture des résultats, a encouragé des initiatives en faveur de la mesure de la performance et d'une gestion centrée sur les résultats (*results-based management*) qui accroissent la demande d'évaluations (SCT, 2005). L'évaluation est donc une pratique en vogue qui alimente un marché prospère et une industrie florissante (Leeuw, 2005).

Actuellement, rares sont les secteurs d'intervention publique qui échappent à l'évaluation, même si des différences intrasectorielles existent en termes d'adhésion à l'exercice, de développements méthodologiques ou de qualité obtenue. Par exemple, en matière d'intervention policière, l'évaluation des résultats des programmes s'est systématisée à partir des années 1990 (Brodeur, 2003 ; Chemers et Reed, 2005). Ce développement s'explique, entre autres, par l'impératif de reddition des comptes à finalité managériale qui accompagne les processus de modernisation de la gestion publique centrés sur des mécanismes de marché

2. Par exemple: « Le programme a-t-il rejoint le groupe cible visé?, Quelle est la validité de la "théorie" qui sous-tend le programme?, Le programme a-t-il été mis en œuvre adéquatement?, Quelle est l'efficacité du programme?, Quel est le ratio entre les coûts et les effets de ce programme? » (Leeuw, 2005).

(ex.: privatisation, contractualisation, etc.) (Hodge et Coghill, 2007; Leeuw, 2005). L'évaluation est parfois présentée comme le moyen le plus sûr de s'assurer qu'«on en a pour son argent», c'est-à-dire que les deniers publics sont alloués aux programmes qui réussissent, ou de favoriser l'apprentissage organisationnel (Leeuw, 2005). L'évaluation semble alimenter naturellement le paradigme managérial de l'État contemporain, même si toutes les évaluations ne s'inscrivent pas dans cet «idéal» puisque certains auteurs présentent l'évaluation comme un instrument de développement de la justice sociale (Lincoln, 2003; Mertens, 2003).

Diversité des pratiques et des conceptions évaluatives

L'évaluation s'est systématisée et structurée, à travers des dispositifs institutionnels et des organisations professionnelles, au cours de la seconde moitié du xx^e siècle. L'histoire de l'évaluation est jalonnée de débats passionnés et se caractérise par un éclectisme affirmé, puisqu'elle n'échappe pas aux considérations et aux réflexions épistémologiques et méthodologiques qui animent plus largement la communauté scientifique (Alkin et Christie, 2004).

De nos jours, l'évaluation regroupe une variété de pratiques très différentes les unes des autres qui gonflent le «courant des études et des données» en vue de documenter le fonctionnement de l'État (Rist et Stame, 2006). Il existe une telle diversité de perspectives qu'il est impossible de les recenser toutes, d'autant plus que les sociétés contemporaines mobilisent abondamment le concept d'évaluation au risque, parfois, de le dénaturer. Le spectre qui s'étend de l'évaluation spontanée à l'évaluation scientifique est très large. Les générations d'évaluation se succèdent dans un mouvement d'accumulation où les tenants de l'évaluation expérimentale côtoient les promoteurs de l'évaluation constructiviste.

Les conceptions relatives à l'évaluation varient d'un pays à l'autre et la notion même d'évaluateur recouvre des réalités très différentes. Les nuances ne se limitent pas seulement à la raison d'être de la profession (un guide, un animateur-facilitateur, un critique, un juge, etc.), mais portent également sur des attributs plus fondamentaux. En effet, le terme évaluateur caractérise un professeur-chercheur, un consultant privé et un agent de l'État. Cette absence de définition claire du rôle de l'évaluateur apparaît aux yeux de certains comme une faiblesse qui affecte l'évaluation en la réduisant à une accumulation de «jugements

subjectifs» ou à une entreprise de *lobbyisme* auprès des décideurs (Roche, 2005).

Les finalités de l'évaluation sont également multiples et dépendent bien souvent des motivations à l'origine de la démarche (Jacob, 2005b). Dans certains cas, la conduite d'une évaluation découle d'une obligation légale et réglementaire, donc d'un rituel auquel doivent se plier les décideurs et les gestionnaires. Dans d'autres cas, l'évaluation a pour objectif de fournir de l'information en vue d'alimenter les processus décisionnels ou d'orienter la conduite de l'action publique. Ainsi, l'évaluation doit concilier les dimensions scientifiques et politiques (Taylor, 2005).

La qualité en débat

Ce pluralisme, qui repose sur des conceptions différentes ou sur une pondération différente de rôle de l'évaluateur, de l'approche méthodologique et de la perception des finalités de l'évaluation (Alkin et Christie, 2004), engendre régulièrement des débats sur la rigueur méthodologique et la validité des connaissances produites. La diversité et l'hétéroclisme qui caractérisent la pratique évaluative apparaissent à certains comme des sources de confusion et des menaces pesant sur l'essor de l'évaluation. Frans Leeuw estime que la crédibilité de la pratique est en danger puisque «n'importe qui peut s'autoproclamer évaluateur et concourir pour l'obtention de contrats. Les commanditaires d'évaluations et de rapports de performance manquent souvent de compétences pour distinguer les évaluateurs professionnels [...] des amateurs bien intentionnés ou des charlatans» (Leeuw, 2005 : 238, trad.).

Plusieurs solutions sont envisagées pour accroître la qualité des évaluations produites (Schwartz et Mayne, 2005). Celles-ci se répartissent entre des pratiques peu contraignantes comme l'adoption d'une charte de qualité ou de standards de bonnes pratiques (ex. : *Joint Committee on Standards for Educational Evaluation*) à des mesures plus contraignantes comme la certification des évaluateurs ou la constitution d'un ordre professionnel en passant par la réalisation de méta-évaluations, c'est-à-dire des évaluations d'évaluation (Stufflebeam, 2001).

C'est dans ce contexte que s'inscrit le débat sur la décision fondée sur des DP en évaluation. Plus largement, la popularité de ce courant se répercute sur l'évaluation puisqu'elle induit des exigences en termes de sophistication de la pratique (Martin et Sanderson, 1999) comme nous le présentons dans la section suivante.

Vers une hiérarchisation des connaissances évaluatives ?

C'est plutôt l'abondance que le manque de données qui constitue de nos jours le principal défi que doivent affronter les décideurs, les gestionnaires et les intervenants. En effet, devant la multiplicité des canaux d'information, il est difficile de déterminer l'ensemble des sources disponibles et d'en apprécier la rigueur et la précision. L'enjeu porte davantage sur la discrimination et la hiérarchisation des connaissances que sur la rareté de données (Marston et Watts, 2003).

La résurgence de l'évaluation expérimentale

Certains auteurs estiment que l'enjeu de la qualité ne sera surmonté que par une hiérarchisation des évaluations produites en s'inspirant de la *Maryland Scientific Methods Scale* ou des standards de la Collaboration Campbell³ qui privilégient les expérimentations (Brodeur, 2003 ; Farrington, 2003 ; Weisburd et Eck, 2004 ; Chemers et Reed, 2005 ; Leeuw, 2005 ; Greenberg *et al.*, 2006). Pour ces auteurs, l'expérimentation est la meilleure source de production de DP car

[...] les expérimentations aléatoires procurent un niveau plus élevé de validité interne que les études non expérimentales. Dans ces expérimentations, les personnes, les lieux ou les organisations sont répartis aléatoirement entre le groupe soumis au traitement et le groupe de contrôle ou de comparaison. Cela signifie que l'on considère que toutes les causes, à l'exception du traitement lui-même, sont distribuées de la même manière entre tous les groupes. En conséquence, si l'effet d'une intervention apparaît, le chercheur peut conclure avec certitude que la cause est l'intervention elle-même et non pas d'autres facteurs confus. (Weisburd et Eck, 2004 : 47, trad.)

Le retour de l'État rationnel s'accompagne, en évaluation, d'une résurgence des méthodologies expérimentales⁴. Au cours des années 1960 et 1970, l'épistémologie expérimentale structure le développement de la pratique évaluative, y compris en matière de sécurité et de justice

3. La Collaboration Campbell est une organisation internationale, fondée en 2000, qui a pour objectif de réaliser des études sur l'efficacité des interventions sociales et comportementales. Voir <http://www.campbellcollaboration.org/>

4. Cette démarche a notamment été employée au Québec pour apprécier les effets d'une opération de prévention des cambriolages (Charest, 2003) et d'un programme de probation intensive du Centre jeunesse de Montréal (Cournoyer et Dionne, 2007).

(Alkin, 2004; Leeuw, 2005), avant de s'effacer au profit d'approches rompant avec les visions positivistes de l'évaluation. L'évaluation expérimentale associe un nombre restreint de participants à la démarche dominée par des conceptions rationalistes (Albaek, 1995) et où existent les distinctions entre les sphères administrative, scientifique et sociale. C'est en réaction à ces considérations que s'élaborent de nouvelles approches⁵ qui entendent démocratiser la production et l'utilisation des connaissances évaluatives à l'ensemble des parties prenantes concernées par le programme à l'étude en vue de mettre l'évaluation au service des plus démunis (Taylor, 2005).

Selon les promoteurs de la résurgence de l'expérimentation et des travaux qui s'inspirent de la Collaboration Campbell, les avantages de cette approche méthodologique sont: la réduction des problèmes de validité des données produites, la meilleure acceptation des résultats dans les milieux de l'éducation, administratifs et politiques, l'accès et la disponibilité des résultats et des devis de recherche, la présence de conditions favorables à l'accumulation de connaissances et de données, la perspective de transfert de connaissances auprès des décideurs et des intervenants, etc. (Leeuw, 2005).

Dès lors, l'expérimentation devient une condition de la qualité de l'évaluation. Les essais aléatoires contrôlés (EAC/*Randomized Controlled Trials*) se multiplient dans tous les secteurs d'intervention publique. Bien souvent, les rapports d'évaluation qui en découlent sont considérés comme une source à intégrer naturellement, voire exclusivement, dans les revues systématiques. À ce sujet, l'affirmation de Lawrence W. Sherman selon laquelle «une revue peut exclure les évaluations qui ne sont pas des EAC, comme c'est le cas en médecine» (2003: 14, trad.) témoigne du fait que les approches expérimentales apparaissent de plus en plus comme des archétypes⁶.

Cette approche évaluative alimente le courant de la décision fondée sur des DP et cherche à se positionner comme l'alternative aux problèmes d'utilisation et de qualité qui caractérisent l'histoire de l'évaluation. Cependant, sans être exemptes d'un préjugé favorable aux devis expé-

5. Il n'existe pas d'appellation uniforme permettant de qualifier ces approches. Le concept d'évaluation démocratique regroupe les évaluations participatives, de quatrième génération, émancipatrices, etc.

6. À ce sujet, l'American Evaluation Association a publié une opinion critique sur la volonté de considérer l'EAC comme la meilleure ou la seule méthode fiable. Voir <http://www.eval.org/doestatement.htm> [consulté le 10 février 2009].

mentaux, les analyses qui comparent les conclusions d'évaluations expérimentales et non expérimentales aboutissent à des résultats contradictoires. Pour les uns, la comparaison montre que les évaluations portant sur le même programme aboutissent à des conclusions différentes en fonction de la méthode retenue (Sherman, 2003) alors que pour d'autres, les résultats sont relativement semblables (Greenberg *et al.*, 2006).

L'utilisation contrainte des connaissances

Derrière la notion de décision fondée sur des données probantes se trouve une sorte d'automaticité dans la prise de décisions et la conduite de l'action publique. Le processus se résume à une affirmation telle que : « Les données nous disent que... et il faut agir en conséquence. » Ce courant cherche à rapprocher les producteurs de connaissances (ex. : les évaluateurs, les chercheurs, les universitaires) des lieux de prise de décisions (Marston et Watts, 2003 : 147). Dans ce contexte, les décideurs, et parfois même les électeurs, sont vus comme des clients que l'évaluateur ou le chercheur doit satisfaire (Shulock, 1999 ; Sherman, 2003). Selon un « modèle rationnel, il est souhaitable de séparer clairement la production de connaissances (par les experts) de l'utilisation de connaissance (par les praticiens) » (Taylor, 2005 : 606, trad.). Il ne s'agit plus seulement de parler aux décideurs mais d'être écouté par ces derniers. En évaluation, la question de l'utilisation des résultats est une préoccupation centrale. Les conclusions et les recommandations d'une évaluation n'engendrent pas systématiquement de modifications sur le terrain. Ainsi, à quelques rares exceptions, l'évaluation n'a qu'une influence limitée auprès des décideurs et des gestionnaires publics (Weiss *et al.*, 2008). Alors que le paradigme managérial accroît la rédaction de rapports (annuels, d'activités ou d'évaluation), il semble que la plupart de ces documents soient lus davantage avant leur publication qu'après (Jacob, 2006).

Les recherches sur l'utilisation des résultats d'évaluation nous obligent à relativiser l'impact que peuvent avoir les évaluateurs et les chercheurs (en sciences sociales) sur les décideurs publics (Shulock, 1999). La question de la pertinence et de la validité des données à mobiliser est au cœur de cette réflexion. À ce sujet, deux visions s'opposent. D'un côté, certains considèrent qu'il existe un stock considérable de connaissances déjà produites, mais qui sont méconnues ou ignorées

par les décideurs. La solution privilégiée pour améliorer la situation est de démocratiser ou de populariser ces connaissances sans chercher à les hiérarchiser et d'assister les utilisateurs potentiels pour qu'ils transposent facilement les connaissances aux particularités des situations auxquelles ils font face (Weiss *et al.*, 2008). D'un autre côté, se trouvent ceux qui estiment qu'il y a des efforts à entreprendre pour promouvoir la décision fondée sur des DP. Pour ces derniers, la solution permettant de surmonter l'inutilisation des connaissances réside dans l'amélioration de la qualité des connaissances produites et la formation des décideurs afin qu'ils deviennent des « consommateurs intelligents de DP complexes » (Sanderson, 2002 : 6, trad.). À cette fin, il est nécessaire de privilégier les connaissances « fiables », « valides » et « solides » (Heinrich, 2007). Ce dernier courant gagne en popularité depuis quelques années, y compris auprès des décideurs et au sein de la communauté des évaluateurs. Par exemple, aux États-Unis, l'adoption, en 1993, du *Governance Performance Results Act* (GPRA) et la plus récente initiative baptisée *Program Assessment Rating Tool* (PART) favorisent la production et l'utilisation de DP pour évaluer l'efficacité des programmes fédéraux (Chemers et Reed, 2005). Ces évaluations sont également utilisées lors de la discussion budgétaire puisque l'Office of Management and Budget (OMB) accorde une plus grande importance aux programmes qui documentent leur efficacité à l'aide d'évaluations expérimentales (Heinrich, 2007). En résumé, « le GPRA et l'initiative PART ont alimenté la nécessité de comprendre les concepts évaluatifs et plus encore, de penser en termes évaluatifs » (Chemers et Reed, 2005 : 263, trad.).

Le mouvement qui se dessine se caractérise par une utilisation contrainte des connaissances évaluatives. Le financement d'un programme est conditionné à la démonstration de son efficacité au moyen d'une évaluation rigoureusement scientifique (Weiss *et al.*, 2008). Cette utilisation contrainte est présente, depuis plusieurs années, dans le secteur de l'aide internationale, où les bailleurs de fonds exigent une reddition de comptes de la part des pays bénéficiaires. La tendance actuelle est la généralisation de cette pratique aux domaines de la prévention (ex. : consommation de drogue, comportements sexuels à risque, violence, etc.) ou de tous les programmes publics. Cette situation soulève de nouveaux défis pour les gestionnaires. En effet, dans un contexte où le financement est conditionné à l'utilisation de données scientifiquement valides, comment les agences locales ou les services avec des effectifs restreints peuvent-ils produire ces connaissances eux-mêmes ? En s'appuyant sur

l'évaluation d'un programme de prévention de consommation de drogue dans les écoles américaines, Carol Weiss *et al.* (2008) constatent que la plupart des organisations, y compris les plus grosses, recourent à des listes de programmes dont l'efficacité a été scientifiquement démontrée. Cette « victoire de l'évaluation [...] s'appuie sur le postulat que les données adéquates ont été colligées et réunies et que la liste des programmes certifiés et prometteurs s'appuie sur une interprétation juste et précise des résultats » (Weiss *et al.* : 38, trad.). À travers leurs analyses, ces chercheurs remettent en question ce postulat et montrent que l'évaluation des programmes qui se trouvent sur ces listes n'est pas toujours exempte de lacunes méthodologiques (ex. : les groupes de comparaison portent sur différentes versions du programme, les études sont menées rapidement après la mise en œuvre du programme) ou éthiques (ex. : risque de favoritisme ou conflit d'intérêts lorsque l'évaluation est réalisée par le concepteur du programme ou par un expert qui a conçu un programme alternatif). Ils concluent qu'il s'agit plutôt d'une nouvelle exigence bureaucratique que d'une avancée pour l'utilisation des connaissances évaluatives. De plus en plus souvent, la réflexion sur la production de connaissances nécessaires à la prise de décisions et l'analyse des données fournies cèdent leur place à une utilisation mécanique des résultats d'expérimentation ou de revues systématiques de la littérature qui soulèvent de nouveaux défis.

Par exemple, l'accroissement des exigences de reddition de comptes et de gestion de la performance entraîne également des effets pervers au sein des services administratifs. Plusieurs études expliquent que les gestionnaires et les intervenants adaptent leur travail en fonction des critères de performance. Il arrive que l'attention porte davantage sur les moyens à mettre en œuvre pour « apparaître performant » que sur la satisfaction des besoins des usagers. Le trucage des instruments de mesure et la manipulation de données prolifèrent à mesure que se développent les instruments de contrôle de la performance (van Thiel et Leeuw, 2002 ; Heinrich, 2007).

Les défis de la rationalisation publique

Comme nous l'avons mentionné à plusieurs reprises dans le présent article, le courant qui met l'accent sur l'utilisation de DP est une « nouvelle » forme de rationalisation de la gestion publique qui n'est pas exempte de limites. Dans cette dernière section, nous recensons quelques

critiques générales formulées dans la littérature sur le sujet. L'espace dont nous disposons ne nous permet pas une énumération exhaustive. Afin d'ouvrir le débat, nous privilégions les enjeux généraux mis en évidence par le retour de l'État rationnel.

Gérer les enjeux de pouvoir

La gestion publique centrée sur les DP est un « nouveau » mode de régulation politique (Taylor, 2005) auquel l'évaluation contribue. Ce courant se développe dans un environnement de plus en plus désidéologisé, c'est-à-dire où les DP priment sur les schèmes politiques. Cette tendance à la dépolitisation est concomitante à d'autres évolutions de la gestion publique comme le « déplacement de responsabilités de la décision publique et de sa mise en œuvre vers des agences publiques indépendantes » (Walker, 2007 : 236, trad.). Or, l'élaboration et la mise en œuvre de politiques et de programmes demeurent un processus politique au cours duquel des valeurs sont prises en considération (Marston et Watts, 2003 ; Naughton, 2005). Il est illusoire de penser que les données produites sont neutres (Williams, 2002) ou qu'elles ne constituent pas des « munitions », utilisées, à bon ou à mauvais escient, dans le débat politique puisque la politique se caractérise par la confrontation d'arguments (Albaek, 1995 ; Hanberger, 2006). À l'inverse, « l'absence de données empiriques étayant les opinions ne signifie pas pour autant que ces opinions sont erronées » (Chalmers, 2003 : 24, trad.). Il apparaît utopique de penser que les DP permettent à elles seules d'aboutir à la dépolitisation du processus décisionnel.

Les courants scientifico-technocratiques ont tendance à occulter le fait que l'évaluation n'est pas toujours un moyen neutre de production de connaissances. L'évaluation comporte des enjeux de pouvoir qu'il ne faut pas oublier. L'évaluation peut avoir une incidence sur la décision publique, peut engendrer des répercussions dans la gestion administrative des programmes et même susciter une reconsidération ou une réinterprétation des problèmes à l'origine de l'intervention publique. Il s'agit aussi d'un instrument stratégique pour organiser, voire orienter le débat public sur des enjeux d'actualité. Dans ce contexte, l'évaluateur doit faire face à des forces d'inertie qui tentent de maintenir le *statu quo* ou de changement qui promeuvent un programme de réformes. Celles-ci sont animées par des motivations ancrées dans l'idéologie partisane, mais aussi dans les préférences des modes d'intervention des adminis-

trations et des associations professionnelles et par les demandes du public et des groupes de pression ou des associations d'usagers (Davies *et al.*, 1999 ; Weiss *et al.*, 2008). En résumé, le débat actuel sur la production et l'utilisation de DP réactualise les oppositions plus anciennes sur la diversité et la validité des connaissances à travers la « dévaluation de certaines formes et sources de connaissances, comme celle des citoyens ordinaires » (Heinrich, 2007 : 269, trad.).

Surmonter les difficultés méthodologiques

Si les ambitions scientifiques affichées par les promoteurs du courant de la décision fondée sur des DP sont claires, force est de constater que leur application concrète soulève également plusieurs difficultés méthodologiques⁷.

La multiplication des attentes et la compréhension différenciée des enjeux apparaissent très souvent en se rendant sur le terrain et en discutant avec plusieurs acteurs impliqués dans l'élaboration et la mise en œuvre d'un programme (élus, gestionnaires, intervenants, etc.). Les certitudes s'effritent au fur et à mesure que la complexité s'accroît. Par exemple, les différents services ou praticiens ont parfois de la difficulté à identifier l'unité d'analyse pertinente voire à définir précisément les contours d'un programme. Dans ce contexte, il devient difficile de pouvoir saisir l'ensemble des phénomènes et d'en rendre compte finement. Pour produire des DP, il est souvent nécessaire de simplifier.

De nombreux chercheurs et évaluateurs s'accordent sur la difficulté inhérente à vouloir appréhender le réel dans sa globalité. Cette tâche est éminemment complexe et impose bien souvent une combinaison de plusieurs disciplines (Jacob, 2008) ou une simplification analytique. Comme nous l'avons vu, les EAC se concentrent principalement sur les résultats des programmes considérés en vue de déterminer ceux qui réussissent en répondant à la question « Qu'est-ce qui fonctionne ? » Malgré sa popularité croissante aux États-Unis, cette approche n'échappe pas aux critiques, comme nous pouvons l'observer par l'entremise de trois enjeux.

7. Dans la présente section, nous nous limitons aux difficultés méthodologiques qui se présentent dans les pays occidentaux. Cependant, le courant de la décision fondée sur des données probantes souhaite la diffusion de ses manières de faire à l'ensemble des régions du monde. Le lecteur intéressé par les enjeux spécifiques des pays en voie de développement peut consulter l'étude de Bowles *et al.*, 2005.

Tout d'abord, on lui reproche une vision étroite, c'est-à-dire de limiter son observation et ses analyses aux effets des programmes ou aux éléments qui peuvent être mesurés (Williams, 2002). Or, n'importe quelle évaluation doit faire face à la difficile nécessité de contrôler la complexité du monde social et de l'environnement dans lequel le programme est mis en œuvre (le « pourquoi? »). Ce besoin de mettre en contexte l'intervention publique est nécessaire pour interpréter adéquatement les résultats produits et comprendre « pourquoi les programmes (ne) fonctionnent (pas) ». Par exemple, l'absence des résultats attendus n'est pas toujours synonyme d'échec du programme. Elle peut s'expliquer par d'autres raisons comme un déficit de mise en œuvre du programme. Certains estiment que l'évaluation ne peut se limiter aux seuls impacts engendrés, mais qu'il est indispensable de comprendre le processus qui influe sur la conduite d'un programme pour mieux juger de sa pertinence. Dans certains cas, des ajustements dans la réalisation du programme s'avèrent suffisants pour atteindre les effets escomptés. (Oakley *et al.*, 2004). Dans le même ordre d'idées, il a été démontré que les résultats négatifs d'une évaluation sont assez souvent camouflés. Les conclusions moins flatteuses pour une organisation ou un gestionnaire sont rarement présentées dans les grandes conférences internationales et ne donnent pas souvent lieu à des publications dans des revues scientifiques. Bref, elles ont moins de chance d'être diffusées et ensuite reprises dans d'autres études y compris les revues systématiques de la littérature (Chalmers, 2003 : 25).

Ensuite, de plus en plus de travaux évaluatifs mettent principalement l'accent sur le groupe cible, c'est-à-dire sur l'acteur qui est considéré comme la source du problème à résoudre (ex. : les gangs de rue, les conducteurs roulant avec les facultés affaiblies, les résidants « peu prévenants » d'un quartier résidentiel). Ce choix (le « qui? ») est critiqué par plusieurs auteurs qui estiment qu'il est réducteur d'appréhender les interventions publiques à l'aune d'un seul acteur, étant donné que les effets des programmes résultent d'interactions entre une variété d'acteurs dont le comportement affecte l'ensemble des résultats (Williams, 2002).

D'autres observateurs estiment que c'est l'objet étudié (le « quoi? ») qui est trop étroit pour pouvoir saisir les changements de comportements induits par les programmes publics (Rossi *et al.*, 2004 ; Chen, 2005). L'intervention publique s'appuie sur une théorie du changement social. La prise en considération de cette théorie est cruciale pour les

promoteurs de l'évaluation réaliste (Pawson, 2002; Pawson et Tilley, 1994). Selon eux, il est important de regarder à l'extérieur de la boîte noire administrative pour comprendre finement le processus qui s'opère lors de la mise en œuvre d'un programme. Or, cette théorie peut être mal formulée ou engendrer des effets de substitution, voire faire plus de mal que de bien (Chalmers, 2003). En ayant une vision trop limitée et étroite de la conduite de l'action publique, il est difficile pour les tenants de l'approche fondée sur des DP de prendre en considération l'inévitable imprévu qui caractérise les interventions de l'État. Or les analystes de politiques savent que le processus de mise en œuvre d'une politique ne se déroule pas toujours comme le prévoient les décideurs (Pressman et Wildavsky, 1984) et qu'il arrive également que les interventions publiques mises en œuvre « adéquatement » produisent des effets inattendus (désirables ou indésirables). Par exemple, la lutte contre les graffitis sur les rames de métro, grâce à un renforcement de la (vidéo)-surveillance et d'une escouade de nettoyage, peut atteindre son objectif tout en déplaçant le problème, c'est-à-dire en contraignant les graphesurs à choisir d'autres lieux moins surveillés. C'est sans doute pour cette raison qu'il est difficile d'aboutir à une vision partagée de ce qui est considéré comme une réussite auprès des gestionnaires des différents secteurs de politiques publiques. La réussite des uns est parfois le début du problème des autres. Le cloisonnement des interventions, la gestion en silo et l'absence d'attention portée à la cohérence des politiques ou des programmes publics affaiblissent la poursuite et l'atteinte de l'intérêt général. Des motivations opérationnelles et réalistes permettent de justifier la déconstruction d'enjeux complexes en problématiques plus ciblées sur le plan de l'intervention. Cependant, à un moment, il est nécessaire de reconsidérer les actions dans une perspective plus large et d'opérer des arbitrages entre des objectifs concurrents. Cette tâche est essentiellement politique et peut s'appuyer sur les connaissances produites sans s'y soumettre, car « aux yeux des ministres, l'idéologie, l'opportunité et les préférences du public entrent en compétition avec les connaissances scientifiques » (Davies *et al.*, 1999 : 4, trad.). En s'inspirant d'un exemple hollandais en matière de prévention de la délinquance, Frans Leeuw *et al.* (2007) estiment que la combinaison d'approches évaluatives qui s'intéressent à la théorie du programme avec les exigences méthodologiques promues par la Collaboration Campbell accroît la crédibilité et l'accessibilité des conclusions et contribue à l'utilisation des résultats par les décideurs.

Enfin, un des défis auxquels doit faire face toute évaluation concerne la définition du moment opportun (le « quand ? ») pour apprécier tous les effets produits (voulus, imprévus, désirables, indésirables) par l'intervention publique. En entamant une évaluation trop tôt, l'évaluateur risque de ne pas pouvoir observer les changements attendus par la mise en œuvre dudit programme sur le terrain et donc de conclure hâtivement à un échec. À l'inverse, en commençant tardivement l'évaluation, il risque de ne plus pouvoir déterminer les effets du programme. Nous comprenons que le moment auquel débute l'évaluation est crucial et qu'il en affecte les résultats. C'est le cas en matière de lutte contre la criminalité où « différentes mesures de prévention peuvent opérer un effet sur le crime avec différents délais » (Charest, 2003 : 34). Il n'existe pas de règle absolue pour déterminer avec précision la durée à partir de laquelle l'ensemble des effets de l'intervention se fera sentir. L'évaluateur doit concilier différents impératifs temporels (annualité du cycle budgétaire, durée de la législature, planification pluriannuelle, etc.) et être conscient des limites de son analyse.

Le paradigme managérial axé sur la mesure de la performance accroît les exigences de suivi à court terme et reconfigure les modes d'action des services. En effet, les gestionnaires privilégient les actions qui induisent des modifications rapidement observables au détriment d'autres initiatives plus lentes à produire des changements. De plus, certains décideurs se trouvent dans une situation où il est difficile d'anticiper les nouveaux défis ou les pressions émergentes qui requièrent des interventions rapides, innovantes et adaptées à des environnements spécifiques. Les attentes formulées par les promoteurs du courant axé sur des DP s'inscrivent dans ce contexte sans véritablement remettre en question, au-delà des traditionnels enjeux de méthode, la validité des résultats produits. Or, comme l'explique Frank Furstenberg (2003), un chercheur qui a étudié les mères célibataires pendant 35 ans, les conclusions varient en fonction des périodes d'analyse. Il estime que les conclusions auxquelles il est parvenu après un suivi de douze ans auraient été très différentes s'il avait interrompu son analyse après cinq ans.

*Être compatible avec les autres « exigences »
de la gestion publique contemporaine*

Le souci de la performance accompagne les processus de modernisation de la gestion publique contemporaine. Les décideurs et les gestionnaires

ne doivent plus seulement faire les bons choix. Ils doivent démontrer aux citoyens (souvent par médias interposés) qu'ils ont pris les bonnes décisions et que leurs actions engendrent une amélioration substantielle de la situation. Dans ce contexte, l'évaluateur apparaît naturellement comme un « expert » (Hanberger, 2006) ou un « producteur de certitudes » (Taylor, 2005 : 605, trad.).

Au-delà des apparences de nouveauté suscitées par les effets de mode, il est intéressant de constater que la décision fondée sur des DP contribue à la réinvention de l'État rationnel (Sanderson, 2004). Il y a quelques décennies, la prise de décisions rationnelle consistait à identifier un problème, à énumérer toutes les options possibles pour y remédier et à rechercher le ratio coûts-bénéfices le plus avantageux (Williams, 2002). Rappelons que ce mode de décision avait été remis en question avec la dénonciation des technocrates et de la gouvernance élitiste. Il s'en est suivie une contestation de l'expertise publique et des décisions imposées d'en haut, sans tenir compte des spécificités des milieux de mise en œuvre. L'évaluation, et plus largement la recherche scientifique, est à nouveau traversée par ce débat puisque, d'un côté, ceux qui mettent l'accent sur la production et l'utilisation de DP estiment qu'il faut s'adresser principalement aux décideurs et aux gestionnaires (Pawson, 2002) tandis que d'autres, à l'instar d'Anders Hanberger (2001), pensent qu'il faut mettre l'évaluation directement au service de la population. Selon ce chercheur, les connaissances évaluatives doivent, avant tout, servir à rendre des comptes et à fournir des informations pour améliorer le débat démocratique et la prise de décisions. Il ne s'agit pas de s'adresser à un seul public, mais de créer des ponts entre les différentes communautés constitutives de la société (Hanberger, 2001 : 212). Dans cette perspective, l'évaluation est un instrument de médiation et de conciliation qui rompt avec les visions managériales ou technocratiques qui animent le courant fondé sur des DP. Celui-ci coexiste avec une demande de participation accrue de la part des citoyens et d'engagement d'acteurs non gouvernementaux dans l'élaboration et la mise en œuvre d'un nombre croissant de programmes publics (OCDE, 2001 ; Williams, 2002). L'évaluation n'est plus uniquement un mouvement descendant mais également un instrument que les citoyens peuvent saisir pour faire remonter de l'information. Ainsi, les dimensions participatives de l'évaluation élargissent la vision « experte » de la production de connaissances à de nouveaux acteurs qui ne sont pas tous jugés légitimes par certains décideurs ou évaluateurs.

De plus, les décideurs et les gestionnaires publics doivent tenir compte d'un impératif éthique croissant. Or, les devis expérimentaux soulèvent des questions éthiques dans leur réalisation. Dans certains cas, selon la nature des interventions expérimentées, un sentiment de favoritisme peut surgir à l'égard d'une partie de la population (Bowles *et al.*, 2005). Dans d'autres cas, certaines expérimentations sont plus hasardeuses et peuvent entraîner des conséquences dommageables sur les citoyens qui y sont soumis. Alors que l'éthique de la recherche se développe pour limiter ces risques et arbitrer les avantages induits en comparaison des inconvénients subis par les participants, les expérimentations ne sont pas systématiquement soumises aux normes et aux principes d'éthique de la recherche⁸.

Enfin, les réformes promues par les théoriciens de la nouvelle gestion publique entraînent la mutation du citoyen-usager en citoyen-client. Ce dernier devient le centre du fonctionnement de l'État et est l'objet d'une attention accrue. C'est dans cet esprit que s'élaborent les enquêtes de satisfaction des usagers, qui produisent également des connaissances qui doivent être intégrées dans la prise de décisions. Or, les attentes de la population peuvent être satisfaites sans pour autant signifier que les interventions sont efficaces au sens convenu par les promoteurs du courant fondé sur les DP. De plus, l'adoption de politiques et de programmes publics doit s'appuyer sur un principe de réalisme en termes de recevabilité des solutions envisagées et de la faisabilité de leur mise en œuvre (Naughton, 2005). Dans ce contexte, quelle donnée doit primer dans la prise de décisions? Qui opérera les arbitrages nécessaires et selon quels principes? Voilà des questions auxquelles il faudra répondre pour concilier les exigences, parfois contradictoires, qui fondent la gestion publique contemporaine.

Conclusion

Comme nous l'avons vu, et pour répondre à la question que nous posions dans l'introduction, certaines formes d'évaluation alimentent le courant fondé sur des DP. La communauté des évaluateurs est telle-

8. Par exemple, au Canada, l'*Énoncé de politique des trois conseils en éthique de la recherche avec des êtres humains* stipule que les évaluations et les démarches d'assurance qualité ne devraient pas être étudiées par un comité d'éthique de la recherche (Secrétariat interagences en éthique de la recherche, 2005).

ment diversifiée qu'elle rassemble les partisans et les adversaires de cette approche. Les différences s'expliquent par les formations des évaluateurs, par les secteurs dans lesquels ils travaillent et par les exigences des commanditaires en ce qui concerne la nature des évaluations qu'ils effectuent. Il n'existe pas de consensus en la matière et comme le démontre l'histoire de la pratique évaluative, celle-ci s'est régulièrement adaptée aux mutations de la gestion publique et s'est souvent développée en s'appuyant sur les arguments de modernisation de l'État.

Au-delà des réflexions théoriques que nous avons présentées dans le présent article, il ne faut pas perdre de vue que l'évaluation est également une pratique utile aux intervenants qui travaillent dans des champs de pratiques très diversifiés (ex. : prévention du crime, services correctionnels, protection de la jeunesse, sécurité publique, etc.). Les connaissances qui découlent de l'évaluation permettent aux intervenants d'appréhender différemment leurs actions et d'en saisir plus précisément les tenants et les aboutissants. En un mot, l'évaluation offre un temps d'arrêt par rapport à « l'urgence du quotidien », qui concourt à une réflexion critique et à un apprentissage individuel et organisationnel qui permet de mettre les interventions en perspective et de les aborder dans leur globalité. Dans chaque champ de pratique, l'on peut dès lors se demander quelle forme de connaissances est la plus appropriée pour alimenter ses réflexions et s'appuyer sur la diversité des pratiques évaluatives que nous avons présentées dans le présent article pour les éclairer.

En conclusion, les approches fondées sur des DP ont le mérite de redynamiser les débats sur les méthodologies employées par l'évaluateur et sur l'utilisation qui est faite de ses conclusions et recommandations par les décideurs, les gestionnaires et les intervenants. C'est en prenant en considération les défis de la rationalisation de l'action publique, que nous venons de présenter, qu'il sera possible d'enrichir les réflexions des défenseurs et des adversaires de ces pratiques et de les reconsidérer à l'aune des besoins et des contraintes des sociétés dans lesquelles elles se développent. Au cours de cette réflexion, il ne faudra pas perdre de vue que, de tout temps, les chercheurs ont eu de la difficulté à trouver leur place dans les relations entre la volonté politique et la gestion administrative (Walker, 2007). Comme l'observait déjà Keynes, « il n'y a rien d'autre que le gouvernement déteste plus que d'être bien informé ; car cela rend le processus décisionnel beaucoup plus compliqué et difficile » (Skidelsky, 1980 : 630, trad.).

Références

- Albaek, 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 V^{ème} République: la construction du «souci de soi de l'État». *Revue française d'administration publique* 102, 307-325.
- Bézès, P. (2007). Construire des bureaucraties webériennes 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 & Londres: 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 (3), 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.
- 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 Millennium* (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 Millennium* (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.
- OCDE (2001). *Des citoyens partenaires: information, consultation, et participation à la formulation des politiques publiques*. Paris: OCDE.
- 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), 291-306.
- 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 & Londres: 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.
- Secrétariat du Conseil du Trésor du Canada (2005). *La santé de la fonction d'évaluation au gouvernement du Canada. Rapport pour l'exercice 2004-2005*. Ottawa: SCT.
- Secrétariat interagences en éthique de la recherche (2005). *Énoncé de politique des trois conseils en éthique de la recherche avec des êtres humains*. Ottawa: Travaux publics et Services gouvernementaux Canada.
- 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.
- 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.

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 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, rationalisation of the decision making process.

RESUMEN • La evaluación de políticas o de programas ha arraigado en las prácticas de la gestión pública contemporánea. En este artículo describimos las múltiples finalidades a que aspira dicha evaluación (por ejemplo generar información, mejorar los programas o nutrir el debate democrático) y los principales debates que animan a la comunidad de evaluadores. Entre éstos, el de la corriente de las decisiones basadas en la evidencia es uno de los más recientes e intensos. Las principales tensiones concierne la jerarquización de los conocimientos evaluativos, la utilización restringida de los conocimientos y el deseo de homogeneizar las prácticas de evaluación. Observando esos desarrollos, constatamos igualmente el resurgimiento de una racionalización en los procesos de toma de decisión y en los principios de gestión pública que parece estar desfasado respecto a las aspiraciones de la población a una mayor participación y transparencia. En este artículo presentamos igualmente los desafíos que enfrentan los responsables de la toma de decisiones y los administradores públicos ante la reinención del Estado racional.

PALABRAS CLAVE • Evaluación de programas, jerarquización de conocimientos, racionalización en la toma de decisión.