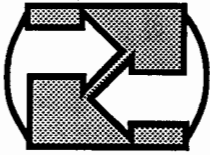


SQEP



SCE

SOCIÉTÉ QUÉBÉCOISE D'ÉVALUATION DE PROGRAMMES,
affiliée à la Société canadienne d'évaluation

Les obstacles à l'évaluation de programmes
dans l'administration publique

Déjeuner-causerie de Réjean Landry,

Université Laval

Québec, 1987

Il n'y a en soi rien de nouveau dans l'idée d'évaluation. Tout jugement porté sur le monde environnant, qu'il s'agisse de notre milieu familial ou de notre univers professionnel, représente une évaluation. Un jugement est une évaluation influençant nos prises de décisions. Tous les programmes gouvernementaux sont élaborés et modifiés sur la base de jugements. La question n'est pas de savoir s'il faut procéder ou non à des évaluations, il s'agit seulement de faire un choix entre différentes méthodes d'évaluation.

A première vue, l'évaluation de programmes représente donc une activité routinière à laquelle s'adonnent quotidiennement les gestionnaires, les législateurs, de même que les consommateurs et les contribuables. En fait, l'évaluation de programmes s'apparente à une activité tellement désirable qu'elle fait l'objet d'attentes très élevées. L'évaluation de programmes n'a toutefois pas réussi à remplir ses promesses. Pire encore, les résultats de la plupart des évaluations de programmes auxquelles procèdent les instances gouvernementales ou les chercheurs ne servent pas lors des prises de décisions concernant la poursuite ou l'abandon des programmes.

En dépit des nombreuses déceptions qu'elle a engendrées, l'évaluation de programmes constitue une industrie relativement chaotique en expansion rapide. Le statut qu'elle revêt aujourd'hui dans le secteur public québécois comporte beaucoup de précarité, aussi bien du côté des structures que de celui des personnes qui y travaillent. L'état déplorable de l'évaluation de programmes au Québec s'explique à la fois par les obstacles à l'évaluation et par des déficiences institutionnelles.

L'institutionnalisation de l'évaluation de programmes

La recherche évaluative ne se ramène pas à une simple question d'application de méthodes. Elle renvoie à des activités de recherche, à des activités de gestion ainsi qu'à des activités politiques qui ne peuvent être comprises adéquatement sans tenir compte de l'évolution des interventions gouvernementales.

L'accroissement du rôle de l'Etat et son corollaire, les programmes publics, constituent des réalités relativement récentes. Les dépenses budgétaires du Québec représentaient 8% du PIB en 1960 contre 25% aujourd'hui. Avant 1960, l'obtention de services d'éducation, de santé et de services sociaux relevait de la responsabilité des individus. Les interventions de l'Etat ont commencé à remplacer temporairement la charité privée au moment de la crise de 1930, puis, lors de la deuxième guerre mondiale. Dans le cas du Québec la création et la croissance des principaux programmes sociaux actuellement en vigueur (c'est-à-dire, ceux qui représentent plus des trois quarts des dépenses budgétaires) remontent aux années soixante.

Les premières grandes tentatives d'évaluation de programmes aux Etats-Unis sont issues de l'implantation des programmes du New Deal et de la seconde guerre mondiale. Les programmes de lutte contre la pauvreté du président Kennedy et ceux de la grande société du président Johnson se sont réalisés dans un contexte mettant l'emphase sur le développement et l'évaluation (au moyen de méthodes expérimentales et quasi expérimentales) d'interventions gouvernementales innovatrices dans de nouveaux domaines. Les jugements plus ou moins favorables portés sur ces programmes ont contribué au cours des années 1970 et

1980, à remettre en cause leur expansion. La remise en cause du rôle de l'Etat et de ses grands programmes publics repose davantage sur des jugements idéologiques que sur des évaluations d'impact, d'efficacité et d'efficience. Par contre, une proportion grandissante de ces critiques idéologiques s'appuient sur des résultats d'études d'évaluation qui mettent en évidence les effets pervers de programmes qui contiennent des incitations qui contribuent à accroître plutôt qu'à réduire les problèmes sociaux. L'avènement de gouvernements à idéologies plus conservatrices ne fait cependant pas disparaître le besoin d'évaluations, il pousse les évaluateurs de programmes à se poser des questions différentes.

Par ailleurs l'activité d'évaluation a commencé à se constituer en profession distincte et à s'institutionnaliser vers le milieu des années soixante-dix. On assiste alors à la création de revues scientifiques et professionnelles, à la publication de nombreux ouvrages ainsi qu'à la naissance des premières associations d'évaluation. De même, les gouvernements de nombreux pays et maintes universités et centres de recherche se sont dotés d'unités d'évaluation ayant une existence institutionnelle officielle. Un bilan récent concernant l'évaluation de programmes soutient que peu de pays ont autant développé leurs institutions d'évaluation que le gouvernement canadien (Hellstern, 1986, 282).

Qu'en est-il au juste dans le cas du Québec? Les premières tentatives d'évaluation ne remontent pas aux interventions gouvernementales de la crise des années trente et de la deuxième guerre mondiale. Les premiers efforts d'évaluation de programmes ont été le fait de commissions royales d'enquête créées par le gouvernement du Québec à partir des années cinquante. A cet

égard, il convient de rappeler les commissions d'enquête sur la constitution (Tremblay), l'agriculture (April), les finances publiques (Salvas), l'éducation (Parent), la santé et le bien-être social (Castonguay-Nepveu). Il n'est pas sans intérêt de rappeler, qu'à la différence des premières grandes tentatives américaines d'évaluation de programmes dominées par les spécialistes des sciences sociales, les premières expériences québécoises d'évaluation ont été le fruit du travail de juges et de juristes jusqu'aux années soixante, moment à partir duquel on assiste à une participation de plus en plus grande des spécialistes des sciences sociales.

Outre les commissions d'enquête gouvernementales, qui ont amené la participation de plusieurs chercheurs universitaires, la seule grande tentative d'évaluation de programmes de type quasi expérimental réalisée par des chercheurs universitaires est l'étude ASOPE (aspirations scolaires et professionnelles des élèves). L'arrêt prématuré de cette grande recherche évaluative s'explique en partie par l'état relatif de sous-développement dans lequel se trouvaient les sciences sociales dans le domaine de la recherche quantitative et l'utilisation de l'informatique. C'est dire que les universités québécoises ne formaient pas beaucoup de spécialistes en évaluation de programmes. Cette situation n'a pas beaucoup évolué depuis dix ans.

Cette carence de diplômés universitaires formés en évaluation de programmes a exercé des effets très importants sur la professionnalisation et l'institutionnalisation de l'évaluation de programmes en milieu gouvernemental. Les évaluateurs de programmes ont dû apprendre sur le tas, dans un environnement où il n'existe pas de critères de qualité des évaluations, ni d'assises institutionnelles solides. Pire encore, l'expertise acquise à l'école de la

pratique tend à se perdre rapidement parce que la carrière professionnelle de l'évaluateur de programmes est souvent courte et précaire dans un milieu où l'évaluation est peu institutionnalisée. Les ministères et organismes du secteur public québécois dotés d'unités d'évaluation sont encore très rares. De plus, le statut institutionnel de l'évaluation de programmes est si précaire qu'une partie des postes d'évaluateurs est occupée par des employés occasionnels engagés pour des contrats de très courte durée. Il n'y a donc pas lieu de s'étonner lorsqu'on réalise qu'une grande partie du travail d'évaluation ne représente rien d'autre que de la cueillette (non systématique) de données qui ne sert pas à la prise de décision.

Les obstacles à l'évaluation

La rareté d'études d'évaluation de programmes ne s'explique pas seulement par des carences institutionnelles; elle dépend aussi de trois grandes catégories d'obstacles: les obstacles méthodologiques, les obstacles administratifs et les obstacles socio-politiques.

Les obstacles méthodologiques

Les obstacles méthodologiques proviennent essentiellement de l'écart séparant les devis méthodologiques utilisés par les évaluateurs et les devis méthodologiques prescrits par les exigences de la recherche scientifique. Bien qu'il ne soit pas possible de passer en revue de façon détaillée toutes les exigences méthodologiques des diverses méthodes d'évaluation de programmes, il convient de signaler les principaux obstacles qu'on rencontre dans l'utilisa-

tion de quelques méthodes classiques, notamment l'évaluation d'impact, d'efficacité et d'efficience puis l'évaluation de coûts.

Evaluer l'impact d'un programme revient à établir un lien de causalité entre une cause et un effet, c'est-à-dire entre l'intervention d'un programme et les effets qu'il a engendrés. Supposons qu'on dispose d'une ou de plusieurs mesures de résultats représentant les objectifs du programme. Tout devis méthodologique doit alors établir une distinction entre les résultats nets et les résultats bruts. Les résultats bruts mesurent tous les changements qui se sont produits subséquentment à l'application du programme: ils renvoient ordinairement à la différence observable entre la valeur des indicateurs de résultats avant l'application du programme et la valeur des mêmes indicateurs après l'application du programme. Les effets nets ou résultats nets renvoient aux résultats uniquement attribuables à un programme en tenant compte des effets de facteurs exogènes au programme. On peut représenter symboliquement la relation entre les résultats bruts et les résultats nets de la façon suivante:

$$\begin{array}{rclclcl}
\text{Résultats} & & \text{Résultats} & & \text{Résultats} & & \text{effets} \\
\text{bruts} & = & \text{nets} & + & \text{de facteurs} & + & \text{du devis} \\
& & \text{(effets de} & & \text{exogènes au} & & \text{méthodologique} \\
& & \text{l'interven-} & & \text{programme} & & \\
& & \text{tion)} & & & &
\end{array}$$

La partie la plus difficile dans la réalisation d'une évaluation d'impact concerne l'identification et la mesure de la fraction des résultats attribuables à des facteurs exogènes au programme évalué. Cette tâche est

d'autant plus difficile que ces facteurs, aussi appelés facteurs de contamination, varient énormément d'un programme à l'autre. Certains de ces facteurs méritent mention parce qu'ils affectent la plupart des programmes. Ce sont:

- les changements endogènes;
- les tendances séculaires;
- les événements d'interférence;
- les tendances de maturation;
- l'absence de contrôle sur la sélection des participants à un programme.

Les effets attribuables au devis méthodologique concernent:

- les effets stochastiques (probabilité de faire des erreurs de jugement dans l'attribution des effets);
- le degré de fiabilité des mesures de résultats;
- le degré de validité des mesures de résultats;
- l'interprétation des valeurs manquantes;
- la représentativité de l'échantillon utilisé.

Bien peu d'évaluateurs disposent du temps et des ressources nécessaires à l'élimination des effets des facteurs exogènes aux interventions des programmes. Il convient toutefois de distinguer trois devis méthodologiques classiques qui se distinguent par la validité de leurs résultats.

Le devis mesurant des résultats bruts:

Résultats bruts (RB) = valeur des mesures de résultats sur les cibles exposées à l'intervention d'un programme (E_2) moins valeur des mesures avant l'intervention (E_1)

$$RB = E_2 - E_1$$

Le devis mesurant des résultats quasi nets:

Résultats quasi nets (RQN) = valeur des mesures de résultats sur les cibles exposées à l'intervention (E_2) moins valeur des mesures de résultats sur un groupe de contrôle équivalent non exposé à l'intervention (C_2)

$$RQN = E_2 - C_2$$

Le devis mesurant des effets nets:

Résultats nets (RN) = différence dans les mesures de résultats avant et après l'exposition des cibles à l'intervention ($E_2 - E_1$) moins différence dans les mesures de résultats aux deux mêmes moments pour le groupe de contrôle ($C_2 - C_1$)

$$RN = (E_2 - E_1) - (C_2 - C_1)$$

Le temps et les ressources à la disposition des évaluateurs de programmes ne permettent généralement pas de prendre plusieurs mesures dans le temps et de constituer des groupes de contrôle. En outre, la constitution de groupes de contrôle peut s'avérer impossible ou bien immorale et injuste. Le droit à

l'égalité des citoyens devant la loi implique que l'intervention au moyen de lois et de règlements fait l'objet d'une application à grande échelle ou d'une participation volontaire. Il est difficile d'imaginer un programme d'aide sociale répartissant aléatoirement des prestations entre un groupe expérimental et un groupe de contrôle.

Par contre, le recours à la mesure des effets bruts donne des résultats non valides. L'évaluation du programme de soutien à l'emploi scientifique réalisée en mars 1986 représente une bonne illustration des aberrations qu'engendre le devis $RB = E_2 - E_1$. Le rapport de cette évaluation conclut, entre autres, que le programme a eu pour effet de:

- faire doubler ou tripler les dépenses de R-D dans les entreprises ayant bénéficié du programme;
- développer 94 nouveaux procédés et 300 nouveaux produits;
- d'augmenter les ventes de 39% (phases I à III), 14% (phase IV) et 14% (phase V).

Les résultats de cette étude indiquent que l'engagement d'un ou de quelques scientifiques et ingénieurs jeunes et sans expérience (mais dynamiques) subventionnés en partie par le gouvernement engendre des effets tellement désirables qu'il conviendrait de recommander au gouvernement d'étendre ce programme à toutes les entreprises du Québec. L'ampleur des effets positifs que cette étude évaluative attribue à un petit programme gouvernemental est si remarquable qu'il est impossible de ne pas avoir des doutes sur la validité des renseignements qu'elle contient. C'est là le cas typique d'une étude qui repose implicitement sur l'évaluation d'effets bruts plutôt que d'effets nets.

Les résultats de telles études ne comportent pas un degré de validité suffisant pour être utile à la prise de décision.

Les évaluations d'impact qui mesurent les effets quasi nets et les effets nets sont donc très rares tandis que celles qui mesurent les effets bruts ne sont pas valides et donc inutilisables pour la prise de décision.

Evaluer l'efficacité et l'efficience des programmes peut sembler offrir une alternative à l'évaluation d'impact. Sous cet angle, le devis méthodologique typique de l'évaluation s'appuie sur l'utilisation de divers ratios, tels que:

Ratios d'efficacité: (réel / prévu)	1) $\frac{\text{résultats atteints}}{\text{résultats prévus}}$
	2) $\frac{\text{activités réalisées}}{\text{activités prévues}}$
	3) $\frac{\text{ressources utilisées}}{\text{ressources prévues}}$
Ratios d'efficience: (extrants / intrants)	1) $\frac{\text{activités réalisées}}{\text{ressources utilisées}}$
	2) $\frac{\text{résultats atteints}}{\text{ressources utilisées}}$
	3) $\frac{\text{résultats atteints}}{\text{activités réalisées}}$

Ce devis de recherche évaluative, en apparence beaucoup plus simple à réaliser, soulève néanmoins quelques difficultés pratiques, notamment en ce qui concerne le choix des indicateurs des mesures de résultats. Les programmes gouvernementaux produisent des biens et des services publics, c'est-à-dire, par

définition, des biens et services qui sont très souvent difficilement mesurables et offerts sans restriction à l'ensemble des citoyens (c'est-à-dire que personne ne peut être empêché de consommer un bien ou service public). Le caractère plus ou moins tangible de la plupart des biens et services issus des programmes gouvernementaux signifie que leurs résultats sont à toutes fins pratiques impossibles à mesurer. Les évaluateurs de programmes finissent souvent par se rabattre sur des indicateurs de résultats qui ne mesurent pas les aspects les plus cruciaux tels que les résultats atteints concernant la qualité des services d'éducation, de santé et de services sociaux. L'introduction du PPBS et l'action du Conseil du Trésor n'ont pas suffi à résoudre cette difficulté à laquelle est confrontée toute évaluation d'efficacité et d'efficience. Ces déficiences du côté des indicateurs des mesures de résultats ont souvent pour conséquence de transformer les évaluations d'efficacité et d'efficience en documents non pertinents et inutilisables pour la prise de décision.

La volonté de contrôler de façon plus serrée la croissance des dépenses budgétaires a incité les gestionnaires à s'engager dans la réalisation d'études de coûts. Ce devis méthodologique, aussi appelé devis du coût unitaire, s'apparente au devis d'évaluation d'efficacité et d'efficience par sa propension à s'appuyer sur l'utilisation de ratios tels que:

Ratios de coût unitaire: (intrants / extrants)	1) $\frac{\text{ressources utilisées}}{\text{activités réalisées}}$	(coûts en ressources par unité d'activité)
	2) $\frac{\text{ressources utilisées}}{\text{résultats atteints}}$	(coûts en ressources par unité de résultats)
	3) $\frac{\text{activités réalisées}}{\text{résultats atteints}}$	(coûts en activités par unité de résultats)

Les devis reposant sur le deuxième et le troisième ratios sont rarement utilisés parce qu'ils soulèvent les mêmes difficultés que les études d'efficacité et d'efficience. D'autre part, le devis articulé autour du premier ratio est très utilisé dans le but de comparer les coûts unitaires entre des établissements, ou encore, entre le Québec et l'Ontario. En apparence assez simple, ce devis n'est pourtant pas toujours facile d'utilisation, car l'identification et la quantification des activités réalisées dans le cadre de l'application d'un programme est une tâche délicate, voire impossible. Il est généralement difficile de séparer totalement les activités d'un programme de celles d'un ministère. Ainsi, par exemple, comment tenir compte des activités du service du personnel ou du service du contentieux d'un ministère dans le cadre de la mesure des activités d'un programme de subventions à l'entreprise.

Les obstacles administratifs

Les obstacles méthodologiques ne sont rien à côté des obstacles administratifs et socio-politiques. Un des obstacles administratifs les plus importants tient à la multiplicité des formes de contrôles en vigueur. L'évaluation de programmes paraît d'autant moins nécessaire qu'il existe un grand nombre de mécanismes sanctionnant les programmes depuis le vote aux élections en passant par les contrôles exercés par l'Assemblée nationale, le Vérificateur général et le Conseil du Trésor. Cette multiplicité de contrôles incite vraisemblablement bon nombre de gestionnaires de programmes à voir l'évaluation comme un contrôle additionnel inutile.

Un autre obstacle administratif découle du cercle vicieux engendré par le faible degré d'institutionnalisation de l'évaluation: comme il se fait peu

d'évaluations de programmes, il n'est pas justifié de créer des unités administratives spécialisées et, en retour, l'absence de telles unités ralentit la croissance des efforts d'évaluation.

Les praticiens de l'évaluation se butent aussi au problème de l'accès inadéquat aux rapports et aux publications concernant l'évaluation de programmes. Les articles publiés dans les revues scientifiques traitent très souvent de raffinements méthodologiques et conceptuels rébarbatifs peu utiles pour le praticien. D'autre part, en étant destinées au sous-ministre du ministère concerné, les évaluations de programmes réalisées par des fonctionnaires et des experts externes peuvent donner lieu à des évaluations indulgentes qui ne constituent pas de bons exemples concernant la façon d'effectuer des évaluations.

Les obstacles socio-politiques

Les obstacles socio-politiques sont les plus difficiles à surmonter. Le plus grand obstacle au développement de l'évaluation tient au fait que ses résultats sont nécessairement dangereux. Quel que soit la performance d'un programme, un rapport d'évaluation révèle toujours des opportunités permettant des améliorations. Les tentatives d'améliorations qui en résultent ne peuvent faire autrement que de mettre en cause de façon plus ou moins significative les responsabilités des gestionnaires des programmes.

Un second obstacle socio-politique qui retarde l'essor de l'évaluation de programmes provient du désaccord profond qui subsiste entre les gestionnaires de programmes, d'une part, et les politiciens et les consommateurs et

contribuables, d'autre part. Les premiers jugent qu'il est plus important d'évaluer l'efficacité et l'efficience que d'évaluer les impacts alors que les seconds penchent très souvent pour l'option inverse. Ce conflit de perspectives contribue lui aussi à limiter les efforts d'évaluation.

Un troisième obstacle socio-politique tient au degré de volatilité des informations recueillies pour effectuer l'évaluation de programmes. Les gestionnaires et les politiciens sont très sensibles aux incidents non routiniers tels que des grèves et des scandales alors que les informations recueillies pour fins d'évaluation de programmes sont orientées vers la gestion routinière. Les résultats de l'évaluation de programmes ne sont pas très utiles pour gérer des situations de crise.

La valeur des informations recueillies pour procéder à des évaluations de programmes peut se dégrader de façon plus lente mais néanmoins significative. Le cas des programmes concernant la délinquance des jeunes fournit une illustration de cette situation. Depuis dix ans, l'interprétation de ce phénomène a évolué assez radicalement en postulant que la délinquance dépend de la société puis, plus récemment, qu'elle dépend de l'individu. Les interventions et les mesures de résultats des programmes diffèrent profondément suivant qu'on adhère à l'une ou l'autre de ces interprétations. L'intervention vise la société plutôt que l'individu dans le premier cas et l'inverse dans le second. Les informations recueillies pour piloter et évaluer les programmes diffèrent alors elles aussi de façon tout aussi significative. Ce genre de changements dans les modèles d'interprétation compliquent singulièrement la mise en place de systèmes de cueillette d'information sur les résultats de programmes.

Création d'incitations susceptibles de surmonter certains obstacles à l'évaluation

Les obstacles méthodologiques, administratifs et socio-politiques dont on vient de faire état ne sont pas totalement insurmontables. Il est possible de créer un système d'incitations qui annule en partie leurs effets tout en stimulant quelque peu le développement de l'évaluation. De façon générale, ces incitations devraient récompenser plutôt que punir les gestionnaires et les ministres qui procèdent à l'évaluation de leurs programmes. Ainsi, par exemple, la peur que suscite l'évaluation disparaîtrait au moins en partie si, en contrepartie d'études évaluatives, les gestionnaires et les politiciens se voyaient accorder un peu plus de flexibilité et d'opportunité d'innover grâce à des primes budgétaires.

D'autre part, l'essor de l'évaluation de programmes serait grandement stimulé par l'émergence d'autres lieux d'évaluation. Ainsi, il convient de se demander dans quelle mesure l'évaluation de programmes devrait revenir aux commissions de l'Assemblée nationale et aux bureaux de recherche des partis politiques. Un premier pas dans cette direction pourrait être fait en modifiant le mandat des commissions parlementaires de façon à ce que le rapport de l'étude détaillée des projets de lois présente un énoncé d'impact rendant obligatoirement compte, par exemple, du nombre de personnes affectées par le projet, de ses impacts économiques et de ses impacts sur la vie privée. Une fois le processus enclenché, les commissions parlementaires pourront utiliser le règlement de l'Assemblée nationale pour se donner des mandats d'initiative parlementaire dans le but de mener à terme des travaux d'enquête de recherche et d'étude sur l'impact, l'efficacité et l'efficience des programmes.

Finalement, il y a lieu d'étendre à l'ensemble des programmes et des instances gouvernementales l'initiative de certains ministères qui font des appels d'offres de soumissions pour l'évaluation de programmes. Ces appels d'offres devraient s'adresser autant aux chercheurs des universités qu'aux bureaux de consultants dans le but d'assurer une pluralité de sources et de modes d'évaluation.

Les obstacles inhérents au processus d'évaluation de programmes sont tellement diversifiés et si complexes qu'il importe de ne pas créer de situation de monopole. L'instauration d'une pluralité de sources et de modes d'évaluation engendrera une situation de compétition qui stimulera l'augmentation du nombre d'évaluation et améliorera leur qualité.