

LE CONTEXTE FRANÇAIS DE L'ÉVALUATION: APPROCHE COMPARATIVE

Bernard Perret
Conseil scientifique de l'évaluation
Paris, France

Résumé: En 1990, le gouvernement français a créé un dispositif institutionnel visant à promouvoir l'évaluation des politiques publiques interministérielles et à favoriser le développement d'une expertise en la matière. Trois ans après le début de cette expérience, l'auteur tente de caractériser les spécificités du modèle français d'évaluation par opposition à ce qu'il nomme le «modèle anglo-saxon» d'évaluation de programme, dominant au niveau international. Outil prometteur de régulation du fonctionnement des appareils publics, l'évaluation de politique s'inscrit moins facilement dans un processus instrumental et contribue moins directement à la préparation des décisions budgétaires.

Abstract: In 1990, the French government created an institutional tool aimed at promoting interdepartmental policy evaluation, and favouring development of expertise in this field. Three years after the start of this experiment, the author attempts to identify the features of the French model of evaluation compared to an Anglo-Saxon model. A promising tool to regulate complex public systems, policy evaluation fits less easily into decision-making processes concerning public resources allocation.

■ Pour analyser le modèle français d'évaluation des politiques publiques, on peut tenter de l'opposer à un modèle anglo-saxon, caractéristique des pays de langue anglaise et dominant au niveau international. Le lecteur voudra bien considérer cette opposition comme une facilité d'exposition destinée à faire ressortir un certain nombre de contrastes entre l'univers de référence implicite de la littérature internationale consacrée à l'évaluation de programme et la réalité des pratiques qui se développent actuellement en France. On analysera successivement les *objets* de l'évaluation, la logique de la *demande d'évaluation*, les *modes d'institutionnalisation et de professionnalisation* de l'évaluation en France et dans quelques pays où l'évaluation a connu un développement plus précoce. Cette étude comparative portera principalement sur les évaluations pratiquées au niveau central de l'administration d'État.

PROGRAMMES ET POLITIQUES PUBLIQUES

Une simple question de terminologie?

Les textes officiels français relatifs à l'évaluation parlent toujours d'*évaluation des politiques publiques*. Les textes de langue anglaise, en revanche, parlent le plus souvent, d'*évaluation de programme*. Le terme de programme désigne, en principe, un ensemble d'actions plus limité et, surtout, défini plus précisément dans ses moyens et ses objectifs opérationnels. Une autre manière d'appréhender la différence entre politiques et programmes est de remarquer que le premier concept se réfère davantage au caractère *normatif* de l'action publique (affirmer des droits, afficher des intentions et des valeurs face à un problème de société), tandis que le second renvoie essentiellement à son moment *instrumental* (la définition précise des moyens et conditions de l'action). En fait, la distinction entre politique et programme n'est pas toujours claire, et l'emploi de l'un ou l'autre terme reflète imparfaitement la réalité: Il est fréquent que l'on parle, en France, d'évaluation de politique publique alors que l'on pratique, en fait, une évaluation de programme. De plus, au niveau régional ou départemental, les évaluations de programme sont plus fréquentes.

De même, l'évaluation de politiques publiques n'est pas absente dans les pays où l'évaluation de programme est le concept dominant. Un rapide tour d'horizon permet de se faire une idée sur le poids respectif des deux notions.

Ainsi aux *États-Unis*, les deux termes sont employés. Cependant dans de nombreux textes, on oppose « policy analysis » et « program evaluation », suggérant que le niveau des politiques doit être l'objet d'une réflexion plus globale à caractère souvent prospectif, tandis que l'essentiel de l'analyse rétrospective devrait porter sur les programmes (Leonard, 1990, p.179). En revanche, un auteur comme Ray C. Rist (1990, p.16) mentionne explicitement le fait que l'évaluation de programme peut être étendue aux politiques:

L'inclusion des politiques au même titre que les programmes vient du fait que beaucoup d'actions et initiatives gouvernementales ne donnent jamais naissance à des programmes au sens conventionnel du terme, c'est à dire, fournir des soins de santé, étendre l'éducation préscolaire, construire des usines de traitement pour les eaux polluées, fournir des facilités pour le stockage des céré-

ales excédentaires, et ainsi de suite. En fait, de nombreuses initiatives stratégiques du gouvernement impliquent des activités telles que la réécriture de la réglementation des banques, le changement des lois fiscales, ou la fixation de nouveaux standards pour la pollution de l'air. Dans chacune de ces circonstances, il n'y a pas de fourniture de services publics, aucune extension ou restriction d'un programme existant, et création d'aucune nouvelle bureaucratie. Chacune de ces mesures représente une procédure administrative gérée par une agence gouvernementale existante. Et cependant, pour chacune d'entre elles, il y a des conséquences que les décideurs voudront connaître. Par exemple: savoir si les politiques ont été correctement appliquées par l'administration et si elles ont eu les effets attendus constituent deux éléments stratégiques de l'information rétrospective.

Contrairement à ce que cette citation pourrait laisser croire, l'opposition programme-politique ne se résume pas à une opposition fourniture d'un service public/réglementation. En France, le terme de politique peut aussi bien désigner un ensemble complexe de programmes qu'une politique réglementaire. C'est dans ce sens que l'on parle, par exemple, d'une « politique de la ville ».

En *Grande-Bretagne*, un document méthodologique rédigé en 1990 par le Trésor britannique s'intitule « policy evaluation » et vise explicitement à la fois les programmes et les politiques « bien que ce guide utilise le terme d'évaluation de politique, les conseils donnés valent également pour les évaluations de programme et de projets ». Le type de démarche préconisé indique que la notion de politique (« les moyens mis en oeuvre pour atteindre les objectifs du gouvernement ») est proche du sens qu'on lui donne en France. Ainsi, la nécessité de reconstruire les objectifs et de s'interroger sur leurs contradictions éventuelles est-elle largement prise en compte: « un objectif de sécurité du département de la santé (HSE) peut être en tension avec le souci de ne pas augmenter les coûts de l'industrie et la complexité des réglementations ». Il n'en reste pas moins que la majeure partie des évaluations portent sur des programmes bien délimités tels que les transferts au secteur économique ou les programmes sociaux. De plus, la majeure partie des textes académiques sur l'évaluation traite de l'évaluation de programme.

Aux *Pays bas*, la « procédure de réexamen » (reconsideration procedure), qui consiste en une remise à plat complète de blocs de dé-

pense publique débouchant sur des décisions budgétaires, est couramment désignée comme une pratique d'évaluation de politiques, le terme de politique désignant dans ce cas un ensemble de programmes concernant un domaine particulier de l'action gouvernementale. En revanche, la charte du Centre d'évaluation des programmes, organisme créé pour développer la recherche et l'enseignement des méthodes scientifiques d'évaluation, indique que le champ de ses activités peut se définir par « la nature programmatique des politiques, mesures, lois ou autres, qui doivent être évaluées: au minimum, une reconstruction en terme d'objectifs et de moyens doit être possible pour que l'évaluation de programme ait un sens ».

Au *Canada*, en *Australie* et en *Allemagne*, on parle presque exclusivement d'évaluation de programme. Un document australien précise que l'évaluation au sein du gouvernement « concerne principalement l'évaluation de programme ou d'éléments de programme, bien qu'elle puisse également inclure le passage en revue de fonctions, de systèmes et d'activités interministérielles » (Auditor General, 1992, p.15).

Conséquences pour les démarches et méthodes d'évaluation: l'évaluation comme confrontation empirique des objectifs, moyens et résultats

Même si la dualité politiques/programmes n'est pas très nette dans la réalité, l'accent différent mis sur l'une ou l'autre notion en France et dans les autres pays où l'évaluation s'est développée révèle des conceptions différentes du management de l'État et explique certaines spécificités des pratiques françaises. La référence spontanée aux politiques plutôt qu'aux programmes renvoie au fait que le débat public se préoccupe davantage de la mise en scène des problèmes, des acteurs et des valeurs qui les animent que des conditions concrètes de l'action.

On peut définir schématiquement un programme par le fait qu'il spécifie un ensemble d'actions à mener par les pouvoirs publics dans le but de produire un service ou d'obtenir une modification de l'état de la société. Les résultats concrets attendus du programme sont précisément définis au départ, de même que sa durée de vie, souvent limitée à quelques années; les crédits budgétaires qui lui sont affectés sont clairement identifiés. Une autre caractéristique importante du programme, surtout dans le contexte anglo-saxon, c'est qu'il relève, normalement, de responsabilités politiques et opérationnelles bien identifiées. En d'autres termes, le programme est fortement structuré par le couple objectifs/moyens.

Une citation illustrera le fait que la notion de programme structure la conception dominante de l'évaluation:

En résumé, les conditions préalables pour évaluer l'efficacité et l'efficacité des programmes comprennent:

1. un programme clairement articulé;
2. les objectifs et les effets attendus clairement précisés;
3. un lien logique entre le programme et les objectifs attendus. (Rutman, 1992, p. 10)

C'est en fonction de cette vision de l'action des pouvoirs publics qu'ont pris corps les premières idées concernant l'évaluation. La conception la plus naturelle de l'évaluation est celle d'un test scientifique de la validité des théories qui sous-tendent un programme, c'est à dire des hypothèses concernant les relations entre moyens mis en oeuvre et les résultats obtenus. Considéré de ce point de vue, un programme est analogue à un traitement médical: l'évaluer c'est se prononcer empiriquement et, si possible, « scientifiquement » sur son efficacité. On connaît les difficultés techniques qui limitent la portée de cette ambition: en tout état de cause, les programmes auxquels sont réellement confrontés les évaluateurs présentent rarement des caractéristiques adaptées à la mise en oeuvre d'une démarche purement empirique.

Même appliquée aux programmes, l'évaluation s'inscrit dans le fonctionnement des organisations et doit répondre à une pluralité de contraintes et de finalités.

Comme on le sait, le développement de l'évaluation s'est quelque peu éloigné de ce paradigme initial. La représentation du programme comme traitement bien spécifié appliqué à la société n'est pas adéquate en pratique pour concevoir une évaluation socialement utile, et cela pour plusieurs raisons:

- la plupart des programmes évalués portent la trace de compromis entre des logiques divergentes et supportent parfois l'héritage d'une succession de décisions contradictoires, ce qui fait que l'on ne peut isoler des objectifs et une théorie d'action univoques;
- le programme effectivement mis en oeuvre s'écarte parfois notablement de l'intention du décideur central, et l'analyse de cet écart constitue en pratique l'un des enjeux importants de l'évaluation;

- les effets d'une intervention publique sur la société sont si divers qu'il faut opérer des choix rarement évidents parmi toutes les variables observables susceptibles d'être modifiées par cette intervention;
- enfin et surtout, l'expérience prouve qu'une évaluation n'est vraiment utile que si la méthodologie mise en oeuvre tient compte des conditions d'appropriation de ses résultats par le commanditaire (et, le cas échéant, les autres « porteurs d'enjeux »).

L'impossibilité de concevoir les programmes comme des réalités univoques (susceptibles d'une représentation unique, valable pour tous les acteurs) et la diversité des objectifs qui peuvent être fixés à leur évaluation expliquent qu'une pluralité de formes d'évaluation de programme se soient développées, notamment aux États-Unis. Les évaluateurs sont obligés de prendre en compte la variété des contextes dans lesquels ils sont amenés à intervenir:

On peut entreprendre une évaluation pour plusieurs raisons: à des fins de management et d'administration, pour juger de l'opportunité d'une modification du programme, pour identifier les moyens de l'améliorer, ou pour répondre aux exigences d'imputabilité des financeurs. À des fins de politique ou de planification, pour tester des idées innovantes sur les moyens à mettre en oeuvre dans le traitement de certains problèmes sociaux, pour décider d'étendre ou de restreindre un programme, ou pour permettre d'argumenter en faveur d'un programme contre un autre. Enfin, une évaluation peut avoir pour but de tester une théorie scientifique particulière. (Freeman et Rossi, 1993, p. 33)

En particulier, la dimension *formative* de l'évaluation est désormais mieux reconnue. Dans bien des cas, la demande d'évaluation recouvre un besoin d'apprentissage et une demande d'aide au pilotage (monitoring) du programme: « l'évaluation formative prend en compte les mille et un métiers ayant trait à la fourniture d'informations aux responsables du programme en vue d'assurer sa bonne marche » (Herman et al., 1993, p. 16).

Cette diversité des finalités traduit la mise en doute au cours des années 80 d'une conception trop étroitement instrumentale de l'évaluation, qui s'est opérée lorsqu'on a pris conscience du fait que les évaluations étaient rarement suivies d'effet (Freeman et Solomon,

1981). Ceci a conduit à mieux distinguer l'utilité d'une évaluation de sa valeur scientifique:

L'attention s'est alors déplacée vers les conditions de mise en vigueur des actions publiques et vers la conception de programmes intégrant dès le départ les moyens de construire une compréhension commune et, si possible, d'atteindre un consensus sur les prérequis de l'évaluation et sur les stratégies visant à maximiser l'applicabilité des résultats et à accroître la probabilité d'amélioration des programmes. (Leca, 1993, p. 4)

Le concept d'« évaluation de la quatrième génération » formulé par Guba et Lincoln (1989, p. 50) représente l'aboutissement de cette prise de conscience: « L'évaluation de la quatrième génération est une forme d'évaluation dans laquelle les revendications, intérêts et enjeux des différents protagonistes servent de principe régulateur (de base pour déterminer quelle information est requise) ». Au total, le paysage de l'évaluation s'est aujourd'hui considérablement diversifié: la pluralité de finalités de l'évaluation a donné lieu à une pluralité de démarches méthodologiques. Parmi les nombreuses typologies qui ont été proposées, la distinction proposée par Chen entre l'évaluation « normative » (centrée sur la confrontation entre la « structure du programme » et ses propres normes « ce qu'il devrait être ») et l'évaluation « causale » ou « descriptive » (centrée sur l'identification et la mesure des « relations causales » impliquées par le programme — « ce que le programme fait réellement ») est particulièrement riche, dans la mesure où elle met bien en évidence le fait que, pour paraphraser une formule de Jean Leca, l'évaluation *d'une* politique s'inscrit toujours peu ou prou *dans* cette politique. En d'autres termes: faute de pouvoir toujours mesurer de manière univoque et exhaustive les effets d'une politique (ou d'un programme), son évaluation est tributaire de référentiels fournis par le contexte politique ou par les protagonistes.

Une dominante plus étroitement instrumentale

Malgré cette remarquable sophistication conceptuelle et cette diversification des pratiques, l'impression demeure pour le lecteur français de textes en langue anglaise d'une vision de l'évaluation plus étroitement instrumentale que celle qui prévaut dans son pays. Freeman et Rossi (p.34), par exemple, identifient à titre principal trois catégories d'évaluation de programme: « (1) l'analyse reliée à la

conception et la planification d'une intervention; (2) le monitoring de la mise en oeuvre d'un programme; (3) l'appréciation de l'efficacité et de l'efficience d'un programme ». Dans tous les cas, on a une identification claire de la position de l'opération d'évaluation dans le cycle de vie du programme et donc, implicitement, une vision univoque du programme lui-même et du type d'usage qui sera fait de l'évaluation. Cette stricte instrumentalisation de l'évaluation est particulièrement nette en Grande-Bretagne. Pour Sylvie Trosa (1993), trois types d'évaluation sont actuellement pratiqués dans ce pays:

- des évaluations qui portent sur des objets circonscrits et limités, notamment les programmes budgétaires (par exemple les actions d'aide industrielle ou d'aide à l'emploi) : les objectifs sont connus, ainsi que les cibles, les hypothèses d'action et les modalités de mise en oeuvre. (. . .)
- des évaluations qui portent sur des politiques plus complexes mais sous la réserve expresse que l'évaluation comporte une dimension prospective forte, conduisant à des propositions sur des problèmes jugés sensibles. L'évaluation « bilan » faisant le compte des réussites et des échecs passés, voire de leurs causes, est jugée peu utile, l'hypothèse étant faite que les circonstances sociales et politiques étant en évolution constante, une « situation » de politique publique ne saurait se reproduire. (. . .)
- des évaluations dont les finalités d'action sont claires et arrêtées à l'avance par les responsables dans leur principe. Les évaluations ne devraient être lancées qu'après une analyse suffisante des marges de manoeuvre permettant d'infléchir les politiques publiques ainsi que de la volonté des responsables politiques de ce faire.

Conception à laquelle on peut opposer frontalement celle qu'exprime Jacques Donzelot (1993), représentatif d'un courant de la recherche évaluative française: « Le propre de l'évaluation est de mettre en relation les présuppositions de l'action avec l'analyse de ses résultats afin d'argumenter sur le sens d'une politique ».

Les spécificités méthodologiques de l'évaluation de politique

En France, les textes officiels concernant l'évaluation font référence à des politiques publiques et non à des programmes. Ce choix ter-

minologique est cohérent avec la réalité: le niveau d'agrégation de l'action publique perçu comme pertinent par les décideurs et l'opinion conduit à ranger sous la même bannière un ensemble d'activités hétérogènes mises en oeuvre par des acteurs diversifiés, en référence à des objectifs relativement généraux (la lutte contre le chômage, contre la toxicomanie . . .). Bien entendu, ces politiques sont la plupart du temps constituées de segments qui peuvent être assimilés à des programmes, mais les questions relatives à l'orientation et à la mise en oeuvre de l'action publique sont rarement posées et discutées au niveau des programmes.

Par rapport à l'évaluation de programme, l'évaluation de politique publique est fortement déportée du côté de l'évaluation normative: il est en pratique assez difficile de se prononcer globalement sur les effets d'une politique complexe. En revanche, il y a presque toujours beaucoup de choses à dire sur la cohérence interne de cette politique, sur la clarté de ses objectifs, sur sa pertinence par rapport au problème de société qu'elle est censé résoudre, et ainsi de suite. Ce qui ne signifie pas que les tentatives faites pour connaître ses effets soient inutiles, bien au contraire. L'effort d'identification des effets et l'interrogation sur la possibilité de les mesurer a deux types de retombées indirectes:

- il aide à reformuler en terme instrumental les objectifs de la politique, ou à souligner le caractère flou et non opérationnel de ces objectifs;
- il produit des effets formatifs en amenant les décideurs et acteurs de terrain à s'interroger plus concrètement sur les résultats de leur action.

Cependant, à supposer même que des connaissances fiables puissent être produites sur les effets propres de la politique, il faut pour qu'elles soient utilisables les compléter par des informations relatives au processus lui-même (les moyens effectivement mobilisés, la manière dont les acteurs de terrain interprètent et mettent en oeuvre les dispositifs prévus dans les textes, les mécanismes d'action effectifs de ces dispositifs, etc.). De plus, pour qu'un jugement évaluatif puisse être porté, il faut que les effets constatés soient confrontés aux objectifs de la politique, ce qui suppose que ceux-ci, généralement flous et partiellement contradictoires, fassent l'objet d'une lecture normative et soient traduits en critères observables. Plus encore que l'évaluation d'un programme, l'évaluation d'une politique nécessite un travail d'intégration et d'interprétation de connaissances hétérogènes. Pour que ce travail aboutisse à des conclusions crédibles

aux yeux du commanditaire politique de l'évaluation, il doit être mené selon des règles qui lui confèrent une double légitimité, scientifique et politique. Dans le contexte de l'évaluation de programme, l'évaluateur professionnel assume souvent seul le poids de cette légitimation, à charge pour lui de mobiliser une information et de mettre en oeuvre des méthodes crédibles aux yeux des utilisateurs de l'évaluation. Dans le cas des évaluations de politique publique telles qu'elles sont pratiquées en France, cette double légitimité est assurée par une *instance d'évaluation* qui précise le questionnement évaluatif, pilote les études, synthétise et interprète leurs résultats.

Par rapport au tableau esquissé du « modèle anglo-saxon », la principale spécificité est une plus grande indétermination des finalités de l'évaluation. Le mandat donné aux instances d'évaluation est en général large, comme si les décideurs voulaient éviter de se prononcer *ex ante* sur le type de conséquences qu'ils entendent tirer du travail d'évaluation. L'usage de l'évaluation s'inscrit naturellement dans une culture française de la décision qui valorise plus que toute chose la liberté d'action des décideurs, souvent contradictoire avec la transparence de leurs intentions. L'évaluation apparaît ainsi comme une ressource cognitive dont l'usage ne peut être prédéterminé.

Pour résumer cette discussion, on pourrait caractériser les démarches d'évaluation de politiques qui se développent actuellement en France par le fait qu'elles accordent une grande place à la régulation du jeu des acteurs par la connaissance. Elles sont cohérentes avec l'idée que le couple acteurs/problèmes joue un rôle plus décisif dans les processus de réforme et de reformulation des politiques que le couple objectifs/moyens. Si cette hypothèse est exacte, elle dessine une perspective originale pour le développement des pratiques évaluatives. L'analyse des logiques d'émergence de la demande d'évaluation permettront de préciser cette originalité et de souligner les difficultés qu'elle comporte.

LES MÉCANISMES DE LA DEMANDE D'ÉVALUATION

L'évaluation au sein du gouvernement comme élément de la préparation des décisions budgétaires et comme outil de responsabilisation des gestionnaires

Dans les pays anglo-saxons et aux Pays-Bas, l'évaluation au sein du gouvernement est, dans la majorité des cas, explicitement liée à la

préparation des décisions budgétaires, et, plus généralement, à un objectif d'amélioration de l'efficacité de la gestion publique. Selon une récente étude comparative sur l'organisation institutionnelle de l'évaluation, cette tendance serait caractéristique de la « seconde vague » d'institutionnalisation de l'évaluation qui a débuté au cours des années 80:

. . . dans les riches années 60, les pays de la première vague étaient gouvernés par des partis réformistes favorables à l'emploi des méthodes de sciences sociales requises par l'évaluation de politique. Au cours de la seconde vague, les gouvernements conservateurs dominant et ils sont confrontés à une situation économique moins favorable. Ils ont eu davantage tendance à utiliser l'évaluation pour remettre en cause, et souvent pour opérer des coupes dans les programmes existants. (Bemelmans-Videc, 1992)

Dans la logique de cette nouvelle orientation, les acteurs dominants de l'évaluation deviennent les Ministres des finances, les Parlements et les Institutions supérieures de contrôle (Cours des comptes).

Aux *Etats-Unis*, c'est l'Office of Management and Budget (OMB), responsable de la préparation du budget fédéral, qui coordonne et impulse l'évaluation de programme au sein du gouvernement. Le rapport sur le Budget fédéral pour 1993 souligne le rôle de l'évaluation de programme dans la préparation du budget:

Le budget reflète un ensemble de choix concernant la priorité relative des programmes et les besoins de financement. Ces choix sont basés, en partie, sur une information analytique et évaluative concernant les conséquences vraisemblables et les résultats des programmes et des politiques. La loi de renforcement du budget (*Budget Enforcement Act*) de 1990 insiste plus que jamais sur la nécessité de planifier et conduire des évaluations de programme rigoureuses pour aider à atteindre le plus haut degré possible de retour sur investissement au niveau fédéral. (*Budget*, 1993, p. 367)

En *Grande-Bretagne*, le Trésor, dont les attributions couvrent l'ensemble de la politique budgétaire et financière, a un rôle moteur dans le développement de l'évaluation de programme au sein de l'exécutif. Il incite les administrations « dépensières » à évaluer leurs programmes ou politiques, et il exige que toute proposition d'action

nouvelle soit accompagnée de proposition sur la manière dont elle pourrait être évaluée. Le Trésor diffuse des consignes méthodologiques mais il laisse l'entière responsabilité du projet au ministère concerné. En revanche, il se réserve le droit de critiquer les évaluations insuffisantes.

Le modèle *canadien* est assez proche, à cet égard, du modèle britannique: le Conseil du Trésor agit sous forme de directives pour inciter les ministères à évaluer leurs programmes. Ces évaluations ne sont cependant pas strictement orientées vers la décision budgétaire: elles visent également à améliorer la gestion des programmes et l'« imputabilité » de leurs responsables (l'obligation de rendre des comptes en fonction d'objectifs précisément fixés). Un récent document souligne également le rôle que peut jouer l'évaluation de programme pour « établir un environnement propice au changement » (Contrôleur général, 1991, p. 2). Le modèle *australien* n'est pas très différent « l'évaluation a plusieurs fonctions. Elle devrait aider les responsables à tous les niveaux à atteindre leurs objectifs au meilleur coût. Elle devrait aussi être un moyen de responsabilisation pour les performances du programme, dans un contexte de plus grande délégation d'autorité aux gestionnaires » (Auditor General, 1992).

En *Allemagne*, l'objectif de l'évaluation pratiquée au sein de l'administration est d'apprécier l'efficacité et l'efficience des programmes. La dimension « imputabilité » n'est en revanche pas mentionnée dans les textes officiels.

Bien qu'elle diffère fondamentalement du PPBS (planning-programming-budgeting system), par son caractère principalement rétrospectif, l'évaluation de programme en constitue dans bien des cas une descendance naturelle: la préoccupation de rendre plus rationnels les critères d'allocation des fonds publics reste fondamentale. C'est particulièrement vrai au Pays-Bas, où la « procédure de réexamen », comme le PPBS, articule très étroitement évaluation, prévision et programmation budgétaire. Même si le critère de rentabilité n'est pas le seul à être pris en compte (l'amélioration de la qualité des prestations fournies à la clientèle des services publics est, par exemple, une préoccupation de plus en plus fréquemment évoquée), le coût des programmes n'est jamais absent de l'horizon des évaluateurs.

Le rôle du Parlement et de la Cour des comptes

La dimension budgétaire de l'évaluation est indissociable de son rôle dans le jeu des pouvoirs, notamment entre l'exécutif et le législatif.

Comme on le sait, démocratie parlementaire et contrôle de l'usage des fonds publics sont historiquement liés. Plus généralement, une évaluation fortement instrumentalisée se développe d'autant plus naturellement que l'espace public est conçu comme un espace de confrontation entre intérêts constitués.

La situation aux *États-Unis* est bien connue: « La richesse de l'évaluation américaine tient beaucoup aujourd'hui au rôle décisif qu'y joue le congrès et au pluralisme des organismes publics ou privés (fondations, universités, cabinets d'audits, think tanks etc.) qui se sont multipliés sur ce terrain » (Viveret, 1989, p. 47). Le General Accounting Office, organisme rattaché au congrès dont les missions de base sont comparables à celles de la Cour des comptes, a une division spécialisée dans l'évaluation qui emploie une centaine de personnes.

Dans les autres pays, le rôle direct du Parlement est moins net. En revanche, le rôle de l'organisme de contrôle (l'équivalent de la Cour des comptes française), qui est davantage au service du Parlement que de l'exécutif, est souvent décisif dans le développement de l'évaluation.

En *Grande Bretagne*, le National Accounting Office effectue des « audits d'efficience » qui sont une forme d'évaluation des programmes et des services. En outre, les constats du NAO servent de base au rapport du « Comptroller » au parlement, qui comporte des recommandations relatives au développement de l'évaluation et, notamment, aux programmes et politiques qui devraient être évaluées.

Aux *Pays Bas*, la Cour des comptes effectue elle-même certaines évaluations de programmes, mais son action en matière d'évaluation est surtout indirecte: elle effectue des audits des systèmes de gestion des ministères et, à ce titre, réalise des « méta-évaluations », c'est-à-dire des revues critiques des évaluations réalisées dans l'administration.

En *Allemagne*, la relance récente de l'évaluation est due à une initiative de la Cour des comptes: dans un rapport publié en 1989, celle-ci a demandé au gouvernement que soit effectivement mis en oeuvre l'article 7 du code budgétaire allemand rédigé en 1970, lequel stipule que lors de la préparation et de l'exécution du budget, les principes d'économie et d'efficience doivent être « observés ».

Il en est de même en *Australie*, où un rapport de la Cour des comptes (National Audit Office) au Parlement (Auditor General, 1992) est consacré à un « audit d'efficience de la mise en oeuvre de l'évaluation

de programme » dans l'administration. Entre autres recommandations, ce rapport préconise une intensification de l'action du ministère des Finances pour convaincre les ministères dépensiers à procéder eux-mêmes à des évaluations.

Au *Canada*, le Vérificateur général a joué dès les années 60 un rôle incitateur dans le développement de l'évaluation de programme. De plus, les activités propres du vérificateur ont évolué pour se rapprocher de l'évaluation: « Au fil des ans, la notion de vérification intégrée s'est élargie. Elle comprend aujourd'hui des activités qui étaient jadis du domaine (exclusif) des évaluateurs de programme (Turgeon, p. 3).

Le cas français

Les mécanismes d'émergence d'une demande d'évaluation qui viennent d'être décrits jouent un moindre rôle en France.

Le Parlement ne joue pas un grand rôle dans la demande d'évaluation, malgré le développement du suivi d'exécution des lois, des commissions d'enquête, et la création en 1983 de l'Office parlementaire des choix technologiques (Conseil scientifique de l'évaluation, 1992, p. 77). En revanche, le rôle de la Cour des comptes va croissant: elle réalise elle-même certaines évaluations (par exemple, récemment, l'évaluation de la gestion de la politique routière et autoroutière), et elle est présente par l'intermédiaire de ses membres dans bon nombre d'organismes ou d'instances d'évaluation. Elle peut en outre saisir le CIME de demandes d'évaluation. Pourtant, au total, son rôle est encore faible: contrairement à certaines de ses homologues étrangères, elle n'a jamais pris d'initiative globale et cohérente pour inciter au développement de l'évaluation à des fins budgétaires au sein de l'administration.

Au sein de l'administration, l'évaluation n'est pas directement liée à la procédure budgétaire. Il est significatif que le mouvement actuel d'institutionnalisation de l'évaluation ne marque aucune continuité avec la Rationalisation des Choix Budgétaires. Celle-ci a été complètement abandonnée en 1983 (Champsaur, 1990), et l'administration qui était chargée de coordonner sa mise en oeuvre, la Direction de la prévision, a été laissée totalement à l'écart de la récente relance de l'évaluation. Celle-ci a été placée sous la responsabilité administrative du Commissariat général du plan, c'est-à-dire hors de l'orbite du ministère de l'Économie et des Finances. Ce constat doit toutefois être nuancé par le fait que le ministère de l'Économie et le

ministère de Budget sont membres permanents du Comité interministériel de l'évaluation et que des services de ces ministères (Direction du budget, Inspection générale des finances) se sont fortement impliqués dans certaines évaluations. Il n'en demeure pas moins que la logique de la demande d'évaluation qui se manifeste dans les choix du Comité interministériel est, au total, fort peu budgétaire.

Evaluation et modernisation

La relance de l'évaluation en France en 1989 devait s'inscrire dans une stratégie de modernisation de l'État fort différente des politiques libérales qui dominent dans les autres pays dont il est question ici. Le souci de maîtriser la dépense publique et d'augmenter son efficacité n'en était évidemment pas absent, mais était contrebalancé par une vision plus dynamique et plus politique du rôle de l'État et des fonctionnaires. L'un des exposés les plus complets de cette politique est la circulaire du 23 février 1989, qui formule quatre orientations principales: « une politique de relations du travail rénovée; une politique de développement des responsabilités; un devoir d'évaluation des politiques publiques; une politique d'accueil et de service à l'égard des usagers ». Dans ce texte, l'évaluation est mise en rapport d'une part avec la responsabilisation des fonctionnaires (de *tous* les fonctionnaires et non pas simplement des « managers » publics) et avec l'amélioration de la qualité du service public. Dans la même logique, le rapport Viveret qui a précédé la mise en place du dispositif interministériel mettait l'accent sur la dimension « démocratique » de l'évaluation, à savoir sa contribution à l'enrichissement et à la clarification du débat public sur les grandes options de l'action gouvernementale, avec la visée explicite de remédier à la perte de légitimité de l'État. Dans la circulaire de 1989, la dimension managériale de l'évaluation et son lien avec la gestion budgétaire sont à peine évoqués par la nécessité d'évaluer « l'efficacité du service ». Le contraste est saisissant avec les textes équivalents émanant de l'administration britannique.

Pourtant, le développement récent de l'évaluation en Grande Bretagne est également lié à un effort de modernisation de l'État. La comparaison entre les deux démarches est instructive: la notion de responsabilité, qui joue un grand rôle dans l'un et l'autre pays, n'y revêt pas exactement la même signification.

Les Britanniques mettent l'accent sur la notion d'imputabilité (accountability), qui signifie « capacité à rendre des comptes », et qui renvoie aux obligations précises du responsable administratif

vis-à-vis de sa hiérarchie, et à la responsabilité de l'administration vis-à-vis du pouvoir politique et du Parlement. C'est avant tout une responsabilité individuelle liée à une position hiérarchique.

Par opposition, la notion de responsabilité développée dans la circulaire de 1989 renvoie davantage à la mobilisation et à la participation de l'ensemble des fonctionnaires. De plus, il s'agit davantage d'une responsabilité diffuse à l'égard de la société que d'une responsabilité vis-à-vis de la hiérarchie liée à l'exercice d'une tâche précise: « C'est . . . en créant les conditions juridiques et matérielles d'une prise de responsabilités effective par le plus grand nombre d'entre eux, en matière de décision comme d'exécution, que l'on pourra mieux répondre aux aspirations des fonctionnaires et mieux satisfaire les usagers » (*Circulaire du 22 février 1989*). Les fonctionnaires sont considérés comme des acteurs à part entière des progrès que doit accomplir l'État pour mieux répondre aux besoins de la société.

Lors de l'installation du Conseil scientifique de l'évaluation le 13 juillet 1990, le Président de la république a clairement suggéré que l'évaluation pourrait contribuer à cette pédagogie de la responsabilité: « Mieux comprendre ce qu'on fait, à quoi cela sert, la perspective dans laquelle on travaille, dans laquelle on s'implique, participer plus directement à un effort de réflexion, je ne pense pas qu'il y ait meilleur moyen pour renforcer l'éthique et la pratique du service public ».

Des principes à la logique réelle de la demande d'évaluation: l'évaluation comme stratégie de légitimation et de contrôle politique sur une action publique

Comme il a été dit plus haut, la nécessité d'une régulation du jeu des acteurs publics par la connaissance semble être au coeur des dynamiques organisationnelles qui génèrent la demande d'évaluation en France. Cependant, cette logique est demeurée peu explicite et peu organisée: le dispositif institutionnel mis en place pour développer l'évaluation des politiques publiques a fait l'impasse sur les mécanismes concrets d'émergence d'une demande d'évaluation au sein de l'administration. En Grande Bretagne ou au Canada, par exemple, la logique budgétaire de la demande d'évaluation est explicite, et le ministère des Finances est clairement désigné pour prendre lui-même l'initiative ou pour orienter la demande d'évaluation des ministères. Force est de reconnaître que le dispositif français n'a pas permis jusqu'ici de faire émerger une « doctrine d'emploi » de l'évaluation, une orientation claire quant au type de politiques

ou d'actions qui doivent être évaluées. Du coup, le profil des politiques effectivement soumises à évaluation relève davantage du jeu des forces à l'intérieur de l'appareil d'État que d'une orientation cohérente. Plusieurs évaluations (aménagement des rythmes scolaire, insertion par l'économique, lutte contre la toxicomanie, sécurité routière, protection des zones humides . . .) ont clairement eu pour finalité de légitimer et renforcer une administration de mission contrainte d'agir dans le cadre d'une coopération avec des administrations plus puissantes. Dans d'autres cas, l'évaluation apparaît comme un mécanisme d'apparition ou de reformulation d'une politique. Prendre l'initiative d'évaluer ou de faire évaluer une politique autorise en effet à réinterpréter ou à reformuler les objectifs de cette politique, voire à mettre à jour des objectifs restés jusque-là flous, contradictoire ou implicites. Enfin, l'évaluation peut être utilisée comme levier d'une remise en cause de la répartition des compétences entre services administratifs. Tel semble être le cas de l'évaluation des politiques en faveur des adolescents en difficulté. Rares sont finalement les cas qui correspondent au schéma « canonique » de l'évaluation: celui d'une administration responsable d'une politique qui se préoccupe de son efficacité.

Ce constat, sévère à certains égards, ne signifie pas que les évaluations ainsi décidées soient de mauvaise qualité ou sans effets pratiques, mais seulement qu'*elles ne sont vraisemblablement pas prioritaires du point de vue de leur utilité, budgétaire ou autre, par rapport à d'autres évaluations envisageables*. Le dispositif interministériel, par défaut de mécanisme explicite de sélection des questions les plus judicieuses, suscite une demande d'évaluation *sui generis*, qui reflète certains dysfonctionnements au sein de l'appareil d'État.

LE MODE DE PROFESSIONNALISATION DE L'ÉVALUATION

L'évaluateur comme professionnel spécialisé

Dans la plupart des pays où l'évaluation s'est développée, et spécialement aux États-Unis et au Canada, la spécificité de la fonction et du métier d'évaluateur sont reconnues. En témoigne, notamment, l'existence d'associations professionnelles, de revues spécialisées, de standards techniques et déontologiques, et ainsi de suite. Ce processus de professionnalisation est cependant loin d'être parvenu à son terme et se heurte à de nombreuses difficultés. La première étant que, comme le notent Freeman et Rossi (p.33) dans le cas des États-Unis, « les praticiens relèvent d'une grande variété de disciplines

académiques et de professions ». Compte tenu des conceptions actuelles de l'évaluation, qui mettent de plus en plus l'accent sur ses dimensions stratégiques et formatives, le métier d'évaluateur n'est pas facile à définir: il doit à la fois maîtriser les outils techniques de l'évaluation et la complexité du jeu des acteurs, être à la fois un stratège et le garant d'une certaine déontologie. Selon J. Plante (1990, p. 29), l'évaluateur de la « quatrième génération » pourrait se définir comme un « animateur-conseil », un « facilitateur » qui joue un rôle central dans la conduite technique de l'évaluation, tout en laissant au commanditaire et aux partenaires de l'évaluation l'entière responsabilité du questionnement et des conclusions finales. Les problèmes de compétence stratégique et de rigueur déontologique sont fréquemment soulignés dans la littérature anglo-saxonne:

A cause de la nature politique du processus évaluatif et du climat politique dans lequel il est conduit et utilisé, il est impératif que vous, en tant qu'évaluateur, examiniez les circonstances de chaque situation évaluative et décidiez si vous pouvez vous conformer aux attentes politiques sans violer votre propre éthique. (Herman et al., 1993, p. 24)

Le cas français: un pluralisme des légitimités et des paradigmes cognitifs

Dans le contexte français, l'émergence d'un métier spécifique de l'évaluation se heurte au fait que l'évaluation s'est développée de manière pluricentrique, dans le prolongement de pratiques cognitives et administratives fortement structurées et fortement implantées au coeur de l'État. On peut identifier au moins quatre pôles de compétences reconnues qui peuvent revendiquer une légitimité professionnelle dans le champ de l'évaluation: les chercheurs en sciences sociales, les services d'étude et de statistiques internes à l'administration, les corps d'inspection et de contrôle, et les cabinets de conseil. D'une certaine manière, l'existence d'une forte tradition de productions de connaissances sur la société au sein de l'administration française (avec l'INSEE, la Direction de la prévision, le Commissariat du plan), est un obstacle à l'autonomisation de l'évaluation comme nouvelle forme d'alliance entre le savoir et l'administration. La seule voie possible semble donc être celle, plus complexe à mettre en oeuvre, d'un décloisonnement progressif de ces différents systèmes de compétence et de leur mise en synergie au sein de processus fortement institutionnalisés. De ce point de vue, le dispositif interministériel paraît bien adapté aux spécificités de la situation française.

L'instance d'évaluation comme évaluateur collectif

Quand on passe en revue les différentes tâches dévolues à l'évaluateur professionnel dans le contexte anglo-saxon, on s'aperçoit qu'elles recouvrent en grande partie celles de l'instance d'évaluation telle que définie par le Conseil scientifique de l'évaluation: c'est en effet celle-ci qui a pour mission de piloter collectivement l'évaluation, de discuter des choix techniques, d'arbitrer entre les attentes des différents partenaires, d'intégrer les résultats des différentes études, de rédiger le rapport final. Bien entendu, il est souhaitable qu'elle comprenne parmi ses membres une ou plusieurs personnes connaissant bien les démarches et méthodes d'évaluation, mais elle assume collectivement la dimension stratégique et déontologique de l'évaluation. Dans d'autres pays, il est fréquent que soient mis en place des comités de pilotage (*steering committee*) ou de conseil (*advisory group*), mais leur rôle est davantage d'informer et de conseiller l'évaluateur que de prendre des décisions stratégiques sur la conduite de l'évaluation.

Le rôle du Conseil scientifique de l'évaluation

Le Conseil scientifique de l'évaluation est une spécificité française, du moins pour ce qui concerne son rôle de régulateur scientifique et déontologique des évaluations financées par le Fonds national de développement de l'évaluation. Il est à noter que le rôle de l'instance d'évaluation décrit dans le paragraphe précédent est indissociable de l'intervention du Conseil scientifique de l'évaluation: la contrainte que fait peser le CSE en donnant un avis *a posteriori* sur la qualité et l'objectivité de l'évaluation permet d'éviter que les discussions au sein de l'instance n'évoluent vers une simple négociation entre les divers intérêts en présence.

NOTE

Le décret du 22 janvier 1990 organise le développement de l'évaluation au sein du gouvernement. Le dispositif interministériel créé par le décret comprend trois nouvelles instances.

Le Comité interministériel de l'évaluation (CIME) est chargé, au termes du décret, de « coordonner les initiatives gouvernementales en matière d'évaluation des politiques publiques ». À ce titre, il a notamment pour prérogative de choisir les projets d'évaluation qui pourront bénéficier du *Fonds national de développement de l'évaluation* (FNDE), créé également par le décret. Une fois connu le

résultat des évaluations, il délibère des suites qu'il convient de leur donner. Le CIME est présidé par le Premier ministre, il comprend comme membres permanents les ministres de l'économie, du budget, de l'intérieur, de la fonction publique et du plan. En outre, le Conseil économique et social peut faire appel au concours du FNDE dans la limite du cinquième de sa dotation.

Le Conseil scientifique de l'évaluation (CSE) est le garant de l'objectivité des méthodes et la valeur scientifique des résultats des évaluations. En toute indépendance vis-à-vis du gouvernement il émet deux types d'avis sur chaque évaluation financée par le FNDE:

- le premier avis, préalable au choix définitif des projets par le CIME, porte sur les méthodes et conditions de réalisation des évaluations;
- le second avis porte sur la « qualité des travaux effectués » au moment de la remise des conclusions par l'évaluateur. Cet avis est rendu public en même temps que ces conclusions elles-mêmes.

Le CSE est composé de 11 membres, choisis par le Président de la république pour leur compétence en matière d'évaluation ou dans le domaine des sciences économiques, sociales ou administratives. Leur indépendance est garantie par le fait que leur mandat de six ans n'est pas renouvelable. Il est actuellement présidé par Jean Leca, professeur à l'Institut d'études politiques de Paris.

Le Commissariat général du plan prépare les délibérations du Comité interministériel de l'évaluation, veille à la mise en oeuvre des décisions prises et à la publication des résultats des travaux d'évaluation. Il suscite l'offre d'évaluation par les administrations et en informe le Comité interministériel de l'évaluation. En liaison avec le Conseil scientifique de l'évaluation, il anime le développement de l'évaluation dans l'administration.

Premier bilan de la procédure: Une quinzaine d'évaluations ont été lancées selon cette procédure depuis 1990, six sont terminées et ont fait l'objet d'un avis final du Conseil scientifique de l'évaluation: Politique de réhabilitation du logement social, Politiques d'insertion sociale des jeunes en difficulté, Accueil des populations défavorisées dans certains services publics, Aménagement des rythmes de vie de l'enfant, Impact du développement de l'informatique sur l'efficacité de l'administration, Aides au logement.

Les missions générales du Conseil scientifique: Le nouveau dispositif ne se limite pas à l'organisation de la procédure de choix et de réalisation des évaluations choisies par le CIME et financées par le FNDE. L'article 9 du décret du 22 janvier 1990 fixe au Conseil scientifique de l'évaluation une mission plus générale de rassemblement et de diffusion de l'information relative à l'évaluation des politiques publiques. Le décret précise en outre que le CSE peut contribuer « à la formation de spécialistes et au développement de la recherche en matière d'évaluation ».

REFERENCES

- Auditor General. (1992). *Audit report No. 23*.
- Bemelmans-Videc, M.-L. (1992). Institutionalizing evaluation: International perspective. Dans J. Mayne, J. Hudson, R. Conner, & M.-L. Bemelmans-Videc (eds.), *Advancing public policy evaluation*. Amsterdam: North Holland.
- Bureau du contrôleur général. (1991). *Aborder les années 90, Perspectives gouvernementales pour l'évaluation des programmes*. Ottawa: Bureau du contrôleur général.
- Champaur, P. (1990). Les leçons d'une expérience. Dans *Outils, pratiques, institutions pour évaluer les politiques publiques*. Paris: La documentation française.
- Conseil scientifique de l'évaluation. (1992). *L'Évaluation de l'expertise à la responsabilité*. Paris: La documentation française.
- Donzelot, J. (1993). *Rapport d'évaluation de la politique de la ville*. Paris: Delegation interministérielle à la ville.
- Freeman, H., & Rossi, P. (1993). *Evaluation: A systematic approach*. Newbury Park, CA: Sage.
- Freeman, H., & Solomon, M. (1981). Evaluation and the uncertain '80s. *Evaluation Studies Annual*, 6, 2-23.
- Government of the United States. (1993). *Budget of the United States Government, Fiscal Year 1993*. Washington: Government Printing Office.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

- Herman, J.L., Lyons Morris, L., & Taylor Fitz-Gibbon, C. (1993). *Evaluator's handbook*. Newbury Park, CA: Sage.
- Leca, J. (1993). *L'Évaluation en développement, deuxième rapport annuel du Conseil scientifique de l'évaluation*. Paris: La documentation française.
- Leonard, H. (1990). The role of policy analysis and evaluation. Dans *Outils . . . op.cit*
- Office of the Comptroller General. (1991). *Aborder les années 90: Perspectives gouvernementales pour l'évaluation des programmes*. Ottawa.
- Plante, J. (1990). Les principes d'une évaluation socialement utile. Dans *Outils . . . op.cit*
- Rist, R. (1990). *Program evaluation and the management of government*. New Brunswick, NJ: Transaction.
- Rutman, L. (1992). Planification d'une étude évaluative. Compte rendu du *Colloque de l'Institut d'études politiques sur l'évaluation des politiques locales*. Paris.
- Trosa, S. (1994). L'évaluation en Grande Bretagne. Dans *L'Évaluation en développement . . . op. cit.*
- Turgeon, J. (n.d.). Intervention au séminaire de Chantilly.
- Viveret, P. (1989). *L'Évaluation des politiques et des actions publiques*. Paris: La documentation française.