



HAL
open science

L'évaluation des politiques publiques entre enjeu politique et enjeu de méthode

Pierre Lascoumes, Frédéric Varone, Xavier Pons

► **To cite this version:**

Pierre Lascoumes, Frédéric Varone (Dir.). L'évaluation des politiques publiques entre enjeu politique et enjeu de méthode. Pierre Lascoumes; Frédéric Varone. CEVIPOF, 44, pp.123, 2006. hal-01337624

HAL Id: hal-01337624

<https://sciencespo.hal.science/hal-01337624v1>

Submitted on 28 Nov 2016

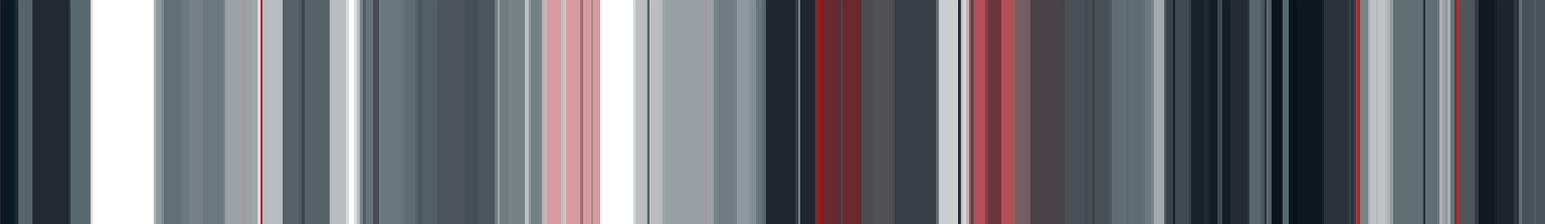
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES CAHIERS DU CEVIPOF

JUIN 2006 / 44


CENTRE DE RECHERCHES POLITIQUES DE SCIENCES PO



L'ÉVALUATION DES POLITIQUES PUBLIQUES ENTRE ENJEU POLITIQUE ET ENJEU DE MÉTHODE

Pierre Lascoumes

Frédéric Varone

en collaboration avec **Xavier Pons**



SCIENCES PO

CNRS
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

L'ÉVALUATION DES POLITIQUES
PUBLIQUES ENTRE ENJEU POLITIQUE
ET ENJEU DE MÉTHODE

Pierre Lascoumes

Frédéric Varone

en collaboration avec **Xavier Pons**

SOMMAIRE

L'IMPOSSIBLE ÉVALUATION DES ACTIONS PUBLIQUES ? PIERRE LASCOURMES (CNRS-CEVIPOF)	7
 PREMIÈRE PARTIE	
BILAN COMPARÉ DE DÉMARCHES	
D'ÉVALUATION DES POLITIQUES PUBLIQUES 25	
INTRODUCTION : L'INSTITUTIONNALISATION DE L'ÉVALUATION DANS UNE PERSPECTIVE COMPARÉE FRÉDÉRIC VARONE (UCL)	27
LA PRATIQUE DE L'ÉVALUATION AU CANADA : LA RECHERCHE DE LA QUALITÉ STEVE JACOB (UNIVERSITÉ DE LAVAL)	33
LES PRATIQUES D'ÉVALUATION DANS LES ORGANISATIONS INTERNATIONALES FLORENT LOISEAU (UNESCO)	37
DÉVELOPPEMENT ET BILAN DES PRATIQUES D'ÉVALUATION AU SEIN DU CANTON DE GENÈVE JEAN-DANIEL DELLEY (UNIVERSITÉ DE GENÈVE)	46
L'ÉVALUATION DES POLITIQUES PUBLIQUES AU COMMISSARIAT GÉNÉRAL AU PLAN. RÉCIT D'UNE EXPÉRIENCE STÉPHANE LE BOULÈRE (COMMISSARIAT GÉNÉRAL DU PLAN)	51
 DEUXIÈME PARTIE	
LES DONNÉES, MÉTHODES ET DÉMARCHES D'ÉVALUATION 59	
ÉLÉMENTS DE RÉFLEXION SUR LA CONSTRUCTION, LES USAGES ET LA PORTÉE DES INDICATEURS STATISTIQUES :	

UNE EXPÉRIENCE PROFESSIONNELLE DANS UNE ADMINISTRATION SOCIALE COMME LA DREES SYLVIE LE MINEZ ET MICHÈLE LELIEVRE (MINISTÈRE DES AFFAIRES SOCIALES, DREES)	65
LA STATISTIQUE PUBLIQUE DE LA NATIONALITÉ ET DES ORIGINES CULTURELLES ET GÉOGRAPHIQUES À L'HEURE DES POLITIQUES PUBLIQUES DE LUTTE CONTRE LES DISCRIMINATIONS JEAN-LUC RICHARD (CRAPE UMR CNRS)	71
L'ANALYSE QUALI-QUANTITATIVE COMPARÉE (AQQC-QCA) : PROMESSES ET PLUS-VALUES POUR L'ÉVALUATION BENOÎT RIHOUX (UCL)	79

TROISIÈME PARTIE

QUELLE PRISE EN COMPTE DES INFORMATIONS

POUR QUELLE HIÉRARCHISATION DES MONDES POSSIBLES ? 87

UNE TYPOLOGIE DES DÉMARCHES D'ÉVALUATION GUILLAUME HOLLARD, MAÎTRE DE CONFÉRENCES EN SCIENCES ÉCONOMIQUES À L'UNIVERSITÉ DE MARNE-LA-VALLÉE (OEP) ET ANTOINE VION, MAÎTRE DE CONFÉRENCES EN SCIENCE POLITIQUE À L'UNIVERSITÉ DE LA MÉDITERRANÉE (LEST)	89
--	----

L'IMPOSSIBLE ÉVALUATION DES ACTIONS PUBLIQUES ?

ENTRE L'ÉPARPILLEMENT
ADMINISTRATIF, LA LOLF
ET UN PROJET D'AUTORITÉ
INDÉPENDANTE...

Pierre LASCOUMES
CNRS-CEVIPOF

Au moment où s'achève la préparation de ce cahier du CEVIPOF, l'évaluation des actions publiques vient en France de connaître un nouvel échec. Il est d'autant plus important, qu'il n'a pas retenu l'attention de beaucoup de commentateurs, comme si une fatalité pesait sur cette activité qui est autant valorisée dans les discours officiels, que consciencieusement évitée dans les pratiques. En adaptant une formule d'un grand spécialiste des décisions publiques Graham Allison, on peut dire « décider sans connaître, c'est choisir les yeux bandés ». Les gouvernants français, depuis plusieurs décennies et quelle que soit leur appartenance politique, semblent tous se conformer sans faille à cet adage alarmant. Qui peut citer UN exemple d'innovation ou de réforme de l'action publique qui ait été préparé ou véritablement accompagné

par une démarche évaluative ? Le suivi de la mise en place du RMI entre 1988 et 1992 demeure la belle exception qui confirme la règle¹.

Le 27 octobre 2005 on apprenait soudainement, à l'occasion d'une conférence de presse du Premier ministre (D. de Villepin), le remplacement du Commissariat au Plan par un Centre d'analyse stratégique directement rattaché à lui. Le commissaire général était démis de ses fonctions. Le seul organisme qui s'efforçait en France, avec plus ou moins de bonheur, de prendre un peu de recul par rapport à l'urgence des décisions et aux pressions constantes de l'agenda politique disparaît. Avec lui, est supprimé le seul lieu où depuis 1990, les autorités publiques centrales avaient amorcé, avec des résultats discutables mais réels, une évaluation des politiques de l'État². Comme le Conseil d'analyse économique ou le Conseil d'analyse de la société, le nouveau Centre d'analyse stratégique devrait être une structure d'expertise gouvernementale, au service des choix du présent en fonction des urgences de l'heure³. S'il fallait un indicateur, non seulement de la faible culture de l'évaluation dans la haute fonction publique française, mais peut-être même de son allergie à ce type de démarche, on pourrait le trouver là⁴. Il est significatif qu'en quinze ans de pratique évaluative institutionnalisée, non seulement celle-ci n'a connu aucun développement véritable mais que son dispositif soit à ce point fragile qu'il soit possible d'en supprimer le cadre sans que personne ne s'en alarme, apparemment.

1 L'article 52 de la loi du 1^{er} décembre 1988 prévoyait une évaluation avant le 2 avril 1992. Celle-ci a été réalisée de façon multi-forme et a permis une adaptation et une poursuite de ce dispositif.

2 Un décret du 22 janvier 1990 avait créé le Conseil national de l'évaluation dont la composition et les mandats avait été modifié le 18 novembre 1998.

3 Alain Etchegoyen, « Le Plan c'était le moyen de réintroduire le long terme dans les décisions publiques », *Le Monde*, 29 octobre 2005.

4 Sur le contexte institutionnel français de l'évaluation et sa comparaison avec d'autres pays, cf. l'excellente recherche de F. Varone et S. Jacob « Institutionnalisation de l'évaluation et nouvelle gestion publique : un état des lieux comparatif », *Revue internationale de politique comparée*, 11 (2), 2004, pp. 271-292 et S. Jacob, F. Varone, Cheminement institutionnel de l'évaluation des politiques publiques en France, Suisse et aux Pays Bas », *Politique et Management public*, 22 (2), 2004, pp. 135-142. Ces deux auteurs ont participé à la journée d'étude dont ce Cahier rend compte.

Pourtant trois grandes raisons au moins auraient dû conduire à un processus inverse, c'est-à-dire au renforcement des activités d'évaluation.

Tout d'abord, depuis la fin du XVIII^e siècle et le développement des « sciences camérales » on sait que la connaissance est une base fondamentale de l'exercice du pouvoir. C'est le savoir sur la société et sur ses besoins qui devrait permettre d'orienter de façon productive l'action publique⁵. Mais paradoxalement, malgré la profusion des instruments d'information (statistiques et autres) la plupart des évaluateurs (individuels ou groupe de) se trouvent régulièrement devant la difficulté de disposer de données adéquates à leur objet. C'est pourquoi une grande part de leur travail reste un travail de collecte et souvent même de production des données nécessaires à leur analyse. Je reviendrai plus loin sur cette question récurrente. De plus, tous les gouvernants redoutent le regard réflexif sur les programmes qu'ils mènent et cherchent régulièrement à instrumentaliser à leur profit les outils et leurs résultats. Autant dire que si beaucoup de démocraties se présentent comme basées sur un « gouvernement rationnel », cela reste du niveau de l'ambition car la relation entre connaissance et action reste très souvent problématique.

Ensuite, depuis le début de la Cinquième République, la question de la modernisation de l'État et celle concomitante des réformes qu'elle exige, sont des leitmotivs des discours des gouvernants⁶. Le Commissariat général du Plan a été pendant vingt ans porteur de cette ambition. Et l'on sait aujourd'hui que son rôle a été davantage de nourrir et de diffuser une culture de la modernité administrative (efficacité, adaptabilité, proximité avec les usagers) que de réellement planifier l'action des gouvernements successifs. Mais cette dynamique d'auto-réforme est très inégalement répartie dans l'administration. Le moment rocardien lié à la volonté de « renouveau du service public » n'a jamais réussi à enrôler véritablement les hiérarchies administratives, et les élites (administrative et politique) françaises ne sont toujours pas formées à la nécessité de rendre compte des raisons de leurs choix et des résultats de leurs actions. C'est pourquoi les changements de politique publique se font plus de façon volontariste et précipitée que sur la base de dossiers convenablement informés et discutés, plus à tâtons que de façon véritablement raisonnée. Les réformes menées récemment sur le

5 J. Gerstlé et al., *Les effets d'information en politique*, Paris, L'Harmattan, 2001, chap. III « Actions publiques », pp. 235-321.

6 Philippe Bézès, « Aux origines des politiques de réforme administrative sous la V^e République : la construction d'un "souci de soi" de l'État », *Revue française d'administration publique*, n° 102, 2003, pp. 307-325.

financement des retraites et de la Sécurité sociale, le CPE ou la non réforme des 35 heures⁷ témoignent du caractère chaotique de la plupart des choix intervenus. Un rapport récent du Sénat note que l'évaluation des politiques publiques a toujours beaucoup de mal à prévaloir sur « les forces profondes qui font que souvent, l'action publique n'est ni vraiment rationnelle, ni vraiment transparente⁸ ».

Enfin, on aurait pu penser que le contexte général de la crise des finances publiques exercerait une pression économique suffisante pour inciter à des stratégies décisionnelles mieux informées. Et qu'ainsi les évaluations de type budgétaire en aval, mais aussi, selon une bonne logique, en amont des mesures publiques se généraliseraient. Ainsi, selon H.-U. Derlien deux des facteurs sur quatre qui ont favorisé le développement de l'évaluation dans plusieurs pays sont le changement de situation fiscale et l'exercice du contrôle financier⁹. Il n'en a rien été pour des raisons diverses, mais en particulier parce que toute organisation répugne à voir mesurer sa dépense par une autorité extérieure (cela fait partie des « secrets maison » bien gardés), et parce que des organismes, en particulier la Cour des comptes et la direction du budget, estiment que cela relève de leur stricte compétence.

L'ÉPARPILLEMENT ADMINISTRATIF

En France, l'évaluation demeure ainsi une activité problématique. Dans un rapport du 30 juin 2004 trois parlementaires soulignent, une fois encore, la faiblesse de cette activité. Certes disent-ils : « Il existe aujourd'hui un large consensus pour reconnaître à l'évaluation des politiques ou des actions publiques le statut d'ardente obligation qui fut, un temps, celui du Plan¹⁰. » Mais

7 Pour l'économiste B. Perret « Nous avons énormément d'éléments précis et souvent contradictoires sur les différents impacts de la réduction du temps de travail, tant du point de vue économique que sociétal. Il suffisait de nommer une commission indépendante, pluraliste, pour établir un véritable bilan de cette loi. Les politiques ne s'y sont pas risqués, ce qui dans un sens n'est pas étonnant », *Le Monde*, 6 juin 2005, Cahier de l'emploi, p. 1

8 J. Bourdin, P. André, J.-P. Plancade, *Rapport d'information sur l'évaluation des politiques publiques en France*, Sénat, rapport n° 392, 30 juin 2004, introduction p. 2.

9 H.-U. Derlien, « Le développement des évaluations dans un contexte international », in W. Bussmann et al. (dir.), *Politiques publiques : évaluation*, Economica, 1998, pp. 5-12.

10 J. Bourdin, P. André, J.-P. Plancade, *op. cit.* note 8

c'est pour ajouter aussitôt « Les efforts effectués pour acclimater l'évaluation des politiques publiques en tant que processus courant de la décision publique sont, quant à eux, souvent restés vains et ont connu l'échec. »

Le rapport attire aussi l'attention sur deux points complémentaires : l'éparpillement des lieux de production de l'évaluation et les lacunes récurrentes d'information sur les politiques publiques.

Les instances publiques qui revendiquent une action évaluative sont nombreuses. Un regard superficiel pourrait laisser croire que contrairement à ce que nous avons indiqué précédemment une culture de l'évaluation a bien été diffusée dans les instances publiques françaises. Mais cette prolifération ne doit pas faire illusion, les activités ainsi qualifiées sont de nature très diverse, leurs résultats très peu coordonnés et leurs effets très limités. Tout d'abord, le Parlement s'est doté de diverses structures destinées à développer sa fonction d'évaluation : la création de deux offices parlementaires d'évaluation des politiques et des lois, la Mission d'évaluation et de contrôle à l'Assemblée nationale, et du Comité d'évaluation des politiques publiques au Sénat. Le Parlement a aussi multiplié ces dernières années les demandes d'évaluation régulière des textes adoptés. Ensuite, la Cour des comptes s'est lancée à côté de ses missions de vérification classique de la régularité des comptes publics dans des actions d'évaluation des politiques. Depuis 1991 la Cour a réalisé 36 « rapports publics particuliers » sur des bilans de l'action publique dans des secteurs précis : de la politique routière et autoroutière en 1992, à celles effectuées sur la politique de la ville en 2002, ou celle consacrée à la politique d'immigration en 2004. Dans ces deux derniers cas les conclusions sont sévères sur le flou des objectifs poursuivis, la superposition de dispositifs sans cohérence et la faible lisibilité des résultats obtenus. Le rapport publié en février 2005 poursuit dans cette voie avec un bilan critique de plusieurs programmes d'action publique dont ceux concernant la prime pour l'emploi et la politique d'immigration.

Enfin, s'inspirant partiellement du rapport Viveret¹¹, un Comité interministériel de l'évaluation (CIME) avait été mis en place le 22 janvier 1990. La lourdeur de la procédure et la faiblesse des moyens n'ont permis la réalisation que de treize évaluations. Réformé le 18 novembre 1998, le Conseil national de l'évaluation (CNE) se substitue au CIME, il est censé être plus autonome dans

11 P. Viveret, *L'évaluation des politiques et des actions publiques*, La Documentation française, 1989.

ses choix et disposer de structures plus resserrées et de procédures allégées. Il a lancé quinze évaluations depuis sa création. Outre la question récurrente du choix des démarches et de leurs coûts (en temps, en budgets) une autre difficulté n'a jamais été surmontée. Celle de la dépendance de ces instances à l'égard des aléas politiques. Ce type de structure ne peut fonctionner qu'avec un intérêt constant et renouvelé des décideurs politiques (qui demeure une des principales faiblesses dans de nombreux pays) afin d'alimenter le processus et permettre le renouvellement des membres motivés et compétents¹². Son avenir est incertain avec la suppression du Commissariat au Plan en octobre 2005.

L'essentiel des activités concrètes d'évaluation française se situe, en fait, à un niveau institutionnel inférieur : celui des ministères, mais là encore il y a des disparités considérables. Globalement, le suivi des politiques en cours et à plus forte raison l'impact des évaluations sur les décisions prises restent faibles. Pour apprécier cela regardons la situation des départements ministériels considérés comme les plus avancés dans le domaine. Ceux qui sont reconnus comme les plus constants dans le recours aux démarches évaluatives et les mieux expérimentés méthodologiquement sont les ministères de l'Équipement, de l'Éducation nationale et celui du Travail et des Affaires sociales. La question centrale de l'information disponible sur les actions et programmes publics n'est pas pour autant résolue. Le rapport parlementaire de 2004 prend ainsi l'exemple de la politique de l'emploi, ensemble de programmes stratégiques s'il en est. Les analyses faites montrent le niveau pléthorique des dispositifs existants et l'incapacité des gouvernements successifs, non seulement à les homogénéiser, mais même à en comprendre l'enchevêtrement malgré des systèmes d'observation très nombreux mais qui sont restés très peu coordonnés. Depuis ces dernières décennies les mesures ont été empilées, juxtaposées sans évaluation préalable et l'avis des parlementaires rejoint le diagnostic de plusieurs spécialistes. Tous s'accordent aujourd'hui pour constater avec l'économiste B. Gazier que l'on est « arrivé à un paradoxe : plus vous faites des politiques de l'emploi, moins vous pouvez les évaluer¹³ ».

12 S. Jacob, « Comparer l'institutionnalisation de l'évaluation des politiques publiques », *Revue française de science politique*, 55 (5-6), 2005, p. 861.

13 B. Gazier, *Tous « sublimes », vers un nouvel plein emploi*, Flammarion, 2003 et dans le même sens P. H. Garraud, *Le chômage et l'action publique, le bricolage institutionnalisé*, L'Harmattan, 2000.

Cet éclatement des lieux administratifs où se pratique, sous des formes diverses, une évaluation, se traduit par une dispersion considérable des données disponibles. La plupart des instances ayant travaillé dans le cadre du Commissariat général du Plan ont relevé qu'elles avaient dû consacrer une part considérable de leur temps de travail à la seule collecte des données retenues comme pertinentes. Il y a non seulement des manques, mais une grande partie des données existantes ont été construites selon des logiques administratives particulières qui ne les rendent pas forcément malléables pour leur intégration dans une démarche évaluative. Les importantes productions de l'INSEE ne fournissent pas le plus souvent les bases adéquates à ce type de travaux tant en raison des catégorisations qu'elles utilisent qu'en raison de leurs conditions particulières d'accès. Comme l'analyse bien A. Desrosières, mesurer n'est pas quantifier : « Dans les sciences sociales, l'emploi immodéré du mot mesurer induit en erreur, en laissant dans l'ombre les conventions de la quantification¹⁴. » La plupart des phénomènes sociaux ne sont appréhendables qu'en fonction d'indices, de données indirectes c'est pourquoi les produits statistiques reconfigurent le monde avant de le décrire. C'est donc bien souvent à un travail préalable, lourd mais essentiel, de démontage des boîtes noires des statistiques existantes que doivent se livrer les évaluateurs soucieux d'utiliser de telles données. C'est pourquoi aussi, ils sont tenus de commencer par produire les informations chiffrées nécessaires à leur démonstration. Le travail récent effectué par P.-M. Menger sur les intermittents du spectacle constitue à ce titre une exception. Il a en effet pu reconstituer (grâce à un travail de reconfiguration important) l'essentiel du système de protection sociale spécifique à ce secteur grâce aux données collectées par les régimes sociaux (cotisations patronales et salariales, prestations de sécurité sociale et de chômage) et éclairer ainsi vingt cinq ans d'échecs des réformes successives¹⁵.

Enfin, une relative ancienneté dans les pratiques évaluatives ne signifie pas pour autant la stabilisation des attentes à l'égard de l'évaluation, ni ne garantit

14 A. Desrosières, « Quantifier », *Genèses*, n° 58, mars 2005, pp. 2-3.

15 P.-M. Menger, *Les intermittents du spectacle, sociologie d'une crise*, EHESS, 2005.

la viabilité à long terme des structures. Un autre exemple d'actualité permet de saisir la fragilité institutionnelle des organes en charge de l'évaluation, et par là même la faible légitimité politique dont ils disposent. En matière de politique éducative le sort du Haut Comité d'évaluation de l'école (HCEE) est significatif des vagues-hésitations qui caractérisent les activités d'évaluation dans beaucoup de secteurs, et le poids des interférences politiques sur ce type d'activité. L'évaluation des politiques d'éducation est prévue depuis la loi d'orientation du 10 juillet 1989 qui assigne à cette activité un triple objectif : faire progresser le système éducatif, diffuser les pratiques innovantes et rendre compte à la Nation. Cette loi confie l'évaluation à l'Inspection générale¹⁶ qui doit procéder à des évaluations à tous les niveaux du système éducatif (départemental, académique, régional et national), en concertation avec les services administratifs compétents, et en s'appuyant notamment sur les productions statistiques de la DEP¹⁷ créée dès 1987. Le décret du 18 juillet 1990 étend cette mission d'évaluation aux inspecteurs territoriaux. Le HCEE est institué par le décret du 16 novembre 2000. Il fait partie au même titre que le CNE¹⁸ et l'Inspection générale de l'administration centrale du ministère de l'Éducation nationale. Il a pour mission d'effectuer un travail d'évaluation de deuxième niveau du système éducatif à partir des travaux effectués par les différents acteurs en charge de l'évaluation que sont les corps d'inspection au sens large, la DEP, mais aussi, dans le cas de l'évaluation des élèves, les enseignants. Le champ de compétence du HCEE couvre aussi bien les questions d'évaluation des acquis des élèves que celles touchant à l'évaluation des performances des établissements ou des pratiques éducatives. Il peut s'intéresser à tous les niveaux du système éducatif, y compris l'enseignement supérieur, voire la formation continue des adultes. Au vu de sa composition tripartite, il devait

¹⁶ Par ce terme, nous désignons le corps d'inspection qui réunit les deux inspections générales que sont l'IGEN (Inspection générale de l'Éducation nationale) et l'IGAEN, devenue en 1999 IGAENR (Inspection générale de l'administration de l'Éducation nationale et de la recherche).

¹⁷ Direction de l'évaluation et de la prospective.

¹⁸ Conseil national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel devenu Comité national d'évaluation.

constituer une sorte de « comité des sages » sur l'école¹⁹. Le HCEE a conduit des évaluations à la fois quantitatives et qualitatives sur des sujets très variés²⁰.

La pratique de son président (Claude Thélot) accompagnait d'une conférence de presse la publication de chaque rapport. Cette publicisation a très tôt été critiquée comme signe d'une indépendance manifeste vis-à-vis des décideurs politiques. Institutionnellement, le HCEE avait été créé pour trancher le débat entre les tenants d'une évaluation externe et purement scientifique (position de C. Allègre à l'époque) et ceux qui défendaient une évaluation interne et participative (position de C. Thélot). Des observateurs ont vu dans cette création un moyen d'écarter ce dernier de ses fonctions à la DEP, car il était alors considéré comme une véritable « voix » de l'école sous un ministre Allègre hostile à toute évaluation interne. À ce moment la DEP est rebaptisée « direction de la programmation et du développement ». La création du HCEE correspondait aussi aux attentes de nombreux évaluateurs (les inspecteurs généraux surtout) qui souhaitaient que leurs rapports aient davantage d'audience interne et d'effets.

Pourtant au bout de quatre ans de fonctionnement, le HCEE est supprimé à son tour par la loi du 23 avril 2005 qui le remplace par le Haut Conseil de l'éducation. Celui-ci doit intégrer d'autres missions à côté des travaux d'étude et d'évaluation. Il doit aussi élaborer le cahier des charges de la formation

19 Le HCEE est composé de trois groupes. Le premier réunit les membres proposés par des organismes représentatifs politiques (Assemblée nationale, Sénat) et professionnels (CFDT, MEDF, SGEN, l'UNEF/FAGE, etc.). Le second comprend des experts français et étrangers (des inspecteurs généraux, des directeurs de recherche, ou des professeurs d'université principalement). Le troisième est constitué par des représentants de l'Éducation nationale ayant voix consultative (chef de service de l'IGAENR, doyen de l'IGEN et directeur de la DEP). Le HCEE se réunit à la demande du Président de la République. Le chef de service de l'IGAENR et le doyen de l'IGEN assistent de droit à toutes les réunions.

20 Dix-neuf rapports furent publiés depuis sa création ce qui représente deux à six rapports par an en moyenne, sur des sujets aussi variés que la réception et l'utilisation en France des évaluations internationales des acquis des élèves, le traitement de la grande difficulté scolaire, les formes et les usages des évaluations d'établissements ou encore l'évaluation des enseignements dans les universités.

dispensée dans les IUFM²¹ et constituer le fameux « socle commun de connaissances et de compétences » de base qu'on est en droit d'attendre de tout élève à la sortie de la scolarité obligatoire. Il intègre pour cela l'ancien Conseil national des programmes. L'évaluation est donc intégrée dans un dispositif plus large. Ces changements à répétition conduisent à s'interroger sur la cohérence à moyen terme de la politique d'évaluation en éducation. Qu'attend-on d'elle ? Qui en décide ? Que fait-on de la connaissance produite ? Les fluctuations observées au mieux fragilisent la production de connaissance sur l'action publique, au pire elles la rendent impossible.

Enfin, les raisons des échecs ne sont pas uniquement externes et structurelles. Par rapport à ce qui s'est passé dans d'autres pays, on peut reprendre le point de vue de Steve Jacob qui estime qu'il y a eu également des problèmes du côté de « l'offre²² ». Pour des raisons intellectuelles autant que pour des logiques d'appartenance institutionnelle, il n'y a jamais eu en France de véritable constitution d'une « communauté des évaluateurs ». Les clivages entre hauts fonctionnaires, universitaires et responsables de bureaux d'étude sont restés très forts. L'opportunité d'ancrer la pratique de l'évaluation dans les pratiques administratives et politiques n'a pas été saisie en raison de divisions et de querelles sur les finalités de l'évaluation. Presque 20 ans après, il n'y a pas de consensus sur le but poursuivi : doit-il être à dominante managériale ou d'apprentissage démocratique²³ ? Le débat est toujours ouvert. Cette situation a freiné le développement des pratiques, la stabilisation des choix méthodologiques et la constitution d'un « corps » d'évaluateurs.

Ce regard plutôt pessimiste sur l'état de l'évaluation en France, surtout au niveau de l'État, est en partie compensé par des innovations et expérimenta-

21 Institut universitaire de formation des maîtres.

22 In *op. cit.* note 12.

23 P. Duran, E. Monnier, « Le développement de l'évaluation en France ; Nécessités techniques et exigences politiques », *Revue française de science politique*, avril 1992, pp. 235-262. J. Leca, « Développement ou crise d'identité de l'évaluation », Conseil scientifique de l'évaluation, *L'Évaluation en développement*, La Documentation française, 1992, pp. 1-12.

tions intéressantes au niveau des Régions (par exemple Bretagne, Rhône-Alpes, Limousin, etc.). La décentralisation (contrats de plan État-Région) et surtout les relations directes nouées entre l'Union européenne et le niveau territorial (Fonds structurels) ont permis le développement de ces activités en faisant de l'évaluation une des conditions de l'attribution des financements. Le colloque ne s'est pas intéressé explicitement à cette dimension, mais elle mériterait à elle seule une synthèse.

Face à l'éparpillement administratif, aux fluctuations de la demande d'évaluation et au faible usage de ses résultats, diverses propositions ont récemment émergées qui se proposent de dépasser ces obstacles. Elles abordent le problème sous deux angles très différents : la LOLF en transformant le cadre d'action budgétaire de l'État veut agir par un contrôle des allocations financières finalisées ; une Autorité indépendante, en chargeant de l'évaluation une agence autonome de la pression de la haute fonction publique et des aléas politiques.

UN NOUVEL INSTRUMENT : LA LOLF

Aujourd'hui beaucoup d'espoirs sont mis dans la LOLF²⁴ qui oblige les administrations à déterminer des objectifs et des programmes par rapport auxquels sont attribués les crédits budgétaires. Si ce texte a pour visée première un renforcement de la portée de l'examen budgétaire par les Assemblées, il accorde une place explicite à l'évaluation des performances en lui donnant une valeur constitutionnelle²⁵ (la LOLF est une loi organique), en élargissant son champ à toutes les actions relevant des finances publiques et en en faisant un complément des activités classiques de contrôle. La LOLF demande aux acteurs publics de déterminer des objectifs d'action, de définir les indicateurs

24 Loi organique relative aux lois de finances du 1^{er} août 2001.

25 L'article 1 de la loi indique : « ... les lois de finances déterminent pour un exercice, la nature, le montant et l'affectation des ressources et des charges de l'État, ainsi que l'équilibre budgétaire et financier qui en résulte. Elles tiennent compte d'un équilibre économique défini, ainsi que des objectifs et des résultats des programmes qu'elles déterminent... »

qui permettent d'en suivre la réalisation et de mettre en place des systèmes d'information finalisés. Les deux lacunes chroniques des politiques publiques françaises devraient ainsi être, sinon résolues, du moins en partie comblées : la détermination des objectifs et une information précise sur les actions et les résultats. Ainsi l'article 51 prévoit que le projet de loi de finances comporte des annexes « accompagnées du projet annuel de performances de chaque programme précisant : la présentation des actions, des coûts associés, des objectifs poursuivis, des résultats obtenus et attendus pour l'année à venir mesurés au moyen d'indicateurs précis dont le choix est justifié. » L'objet n'est pas ici la discussion d'ensemble de ce nouvel instrument de suivi de l'action publique, mais de s'interroger sur sa capacité à intensifier les activités d'évaluation. Et ceci d'autant plus que nous sommes au tout début de la mise en œuvre. Deux éléments incitent cependant à la prudence sur la capacité de la LOLF à parer aux insuffisances des pratiques évaluatives. Il s'agit tout d'abord d'un instrument de comptabilité publique et sa pertinence repose en grande partie sur la validité des nomenclatures qui ont été élaborées. Comme l'indiquent les parlementaires qui ont anticipé une partie des biais possibles : la logique de la LOLF « pourrait s'évaporer si la définition pratique des missions et des programmes devait déboucher sur une nomenclature budgétaire ne se conformant pas à la volonté exprimée par le législateur organique. Tel serait en particulier le cas si une logique institutionnelle devait prévaloir sur une approche fonctionnelle, assimilant mission à ministère et programme à direction administrative²⁶. » On peut élargir l'interrogation à l'ensemble des catégorisations utilisées, quels agrégats recouvrent-elles, quelle homogénéité présentent-elles entre elles et par rapport à celles élaborées par d'autres secteurs administratifs ? Un élément de réponse peut être recherché dans l'analyse des auteurs de ces nomenclatures et celle des procédures qu'ils ont utilisé pour les définir et les valider. On aurait pu penser que l'élaboration des grilles d'activités et des indicateurs de performance offrait une excellente opportunité pour une auto-évaluation des objectifs, moyens et résultats des secteurs administratifs. Un temps d'objectivation des pratiques en cours,

26 Rapport du Sénat n° 388 du 9 juillet 2003 : *Mise en œuvre de la LOLF : un outil au service de la réforme de l'État*, Jean Arthuis, Commission des finances.

d'analyse des activités et de leurs produits aurait pu susciter, dans le même temps, un apprentissage collectif et la négociation de nouveaux objectifs et moyens de suivi. Pour reprendre une terminologie classique, une telle évaluation en amont de la réforme aurait pu être à la fois endo-formative (regard critique sur ce qui est à l'œuvre) et normative (détermination de buts renouvelés). Pour cela, il aurait fallu que la mise en œuvre de la LOLF se fasse « par le bas » (*bottom up*), en s'appuyant d'abord sur la compréhension que les acteurs opérationnels de l'action publique ont de leurs actions et des nouvelles orientations qu'ils peuvent lui donner. Cela aurait exigé autant de volonté politique que de temps pour donner du sens à l'exercice. Il n'en a pas été ainsi, même si le projet de LOLF était en gestation depuis plusieurs années²⁷. Deux réserves principales sont aujourd'hui formulées sur ce nouvel instrument : l'un relatif à ses conditions d'élaboration, l'autre relatif à sa technique propre. D'un côté, l'entrée en vigueur s'est faite à marche forcée et selon une dynamique hiérarchique descendante (*top down*) accompagnée d'une concertation a minima. Dans la quasi-totalité des ministères se sont les services financiers qui ont été en charge de la réalisation des outils de la LOLF et ils ont accompli leur ouvrage de façon assez isolée en utilisant les apports de certains services de façon plutôt instrumentale (fourniture d'indicateurs pour des catégories d'action prédéfinies..). D'un autre côté, la logique d'agrégation pour l'élaboration des objectifs s'est faite davantage à partir de critères organisationnels qu'en fonction de finalités concrètes. Quant aux critères de choix des indicateurs de performance, ils nous informent davantage sur ce qui est aujourd'hui mesurable (on est parti des données déjà existantes) que sur les

27 Le projet trouve son origine en 1998 avec l'initiative de parlementaires, notamment le député socialiste Didier Migaud alors rapporteur général de la commission des finances, dans la création d'un groupe de travail sur l'efficacité de la dépense publique et le contrôle parlementaire. Laurent Fabius, alors président de l'Assemblée nationale, défend l'idée d'un renforcement des pouvoirs de contrôle du Parlement. L'ensemble débouche sur un rapport fondateur (Rapport du groupe de travail sur l'efficacité de la dépense publique et le contrôle parlementaire. Contrôler réellement pour dépenser mieux et prélever moins. Assemblée nationale, 27 janvier 1999). Le retour de L. Fabius au ministère de l'Économie et des Finances, en mars 2000, favorise l'inscription de la réforme à l'agenda autour d'une stratégie d'alliances entre le gouvernement, l'Assemblée nationale, le Sénat et la direction du Budget.

variables les plus pertinentes pour rendre compte des objectifs affichés, mais qui seraient en général à produire²⁸... Au bout du compte, cette rationalisation de la présentation budgétaire ne semble pas susceptible de se substituer à l'évaluation. Elle devrait clarifier les affectations financières, permettre une certaine appréciation de la performance, mais l'instrument ne paraît pas apte à aider à la compréhension des dynamiques d'action et des raisons pratiques qui les gouvernent...

UNE AUTORITÉ ADMINISTRATIVE INDÉPENDANTE ?

Face à l'éparpillement administratif des pratiques actuelles et face aux limites de la LOLF que peut-on proposer pour donner un sens renouvelé aux activités d'évaluation ? Le rapport parlementaire précité voit une solution dans la création d'une nouvelle institution la Haute autorité de l'évaluation (HAE) qui superviserait un ensemble de structures²⁹. La Haute autorité aurait pour tâche principale la définition et le contrôle de « bonnes pratiques de l'évaluation » par la validation de méthodes et leur diffusion auprès des acteurs concernés (via des actions d'information et de formation). Elle serait aussi un superviseur des actions évaluatives menées ; enfin, elle définirait et garantirait les règles déontologiques de fonctionnement des instances. Une seconde mission compléterait de façon plus originale la première, la HAE serait aussi en charge de « favoriser l'efficacité de l'évaluation ». Dans ce sens il lui appartiendrait tout d'abord de veiller au bon déroulement des évaluations tout au long de leur accomplissement (en réglant les tensions entre commanditaire et acteurs de l'évaluation, en débloquent l'accès à des sources d'information, etc.). Ensuite, elle serait en charge de la lisibilité des travaux d'évaluation et du suivi de leurs effets. C'est le caractère très délicat et potentiellement conflictuel de ces missions à accomplir dans le contexte politico-administratif français déjà

28 A. Lacaze, « La LOLF : simple outil de management, ou dogme écrasant ? », *Gérer et comprendre*, *Annales des mines*, n° 81, 2005, pp. 5-13

29 Cette solution rapprocherait en partie la France des situations anglo-saxonnes, le *General Accounting Office* aux États-Unis (créé en 1974), ou le *National Audit Office* (créé en 1983) devenu aujourd'hui l'*Audit Commission* en Grande Bretagne.

évoqué, qui impose selon les rapporteurs le recours à une autorité dotée de la plus grande indépendance. Les opérations d'évaluation elles-mêmes seraient menées à partir de Commissions de l'évaluation créées par grands secteurs de l'action publique et organisées sur un mode pluraliste (sur un modèle proche de l'actuel Conseil national de l'évaluation). Celles-ci sélectionneraient les demandes³⁰ et reformuleraient au besoin les attentes du commanditaire. Elles formeraient des instances ad-hoc en fonction des projets retenus.

Les parlementaires entendent ainsi doter la Haute autorité d'une légitimité particulière qui lui permette de résister aux pressions politiques et aux résistances corporatistes de la haute administration et des grands corps. Ce projet s'inscrit en effet en continuité avec les attributs classiquement attribués aux Autorités administratives indépendantes, en particulier celles relevées par le Conseil d'État en 2001 qui voit en elle des instances dotées d'une crédibilité particulière³¹. Cette dernière se déclinant en termes d'impartialité, de professionnalisme et de souci d'efficacité. Le Conseil en fait aussi un « véritable stimulant de l'action publique » en raison des médiations que les AAI opèrent entre acteurs publics cloisonnés et entre eux et la société civile. Mais prendre ainsi le problème par le haut en ajoutant à l'édifice déjà complexe une couche supplémentaire, est-ce le meilleur moyen de traiter des pesanteurs institutionnelles qui resteraient inchangées ? Sur la base d'un travail de comparaison internationale S. Jacob est très réservé sur l'efficacité de ces organismes centralisés et unificateurs³². Il met plutôt l'accent sur le rôle moteur de réformateurs administratifs et d'entrepreneurs politiques dans la diffusion et l'appropriation des pratiques évaluatives.

³⁰ Les possibilités de saisine seraient très larges, outre les grandes institutions publiques, les organismes professionnels représentatifs et des initiatives populaires de plus de 500 000 signatures pourraient demander l'évaluation de programmes.

³¹ Conseil d'État, *Rapport public pour 2001, Les autorités administratives indépendantes*, La Documentation française, 2002, pp. 253-287. Également, M. Gentot, *Les autorités administratives indépendantes*, Clefs, Montchrestien, 1991, p. 147 s.

³² *Op. cit.* note 4 b.

Deux points de discussion peuvent être introduits sur cette proposition de création d'une AAI en matière d'évaluation. Tout d'abord, on comprend bien dans ce projet le souci d'échapper d'un côté aux intrusions et instrumentalisation politiques, et d'un autre côté à l'emprise technocratique des grands corps et à leur prétention au monopole de l'expertise. Mais comme le montre très nettement la contribution de J.-D. Delley sur les pratiques d'évaluation dans le canton de Genève, la question de la tenue à distance du politique ne peut être une fin en soi. Au contraire, la validité des commandes en amont et la réception des conclusions évaluatives sont inséparables de la qualité des échanges entre évaluateurs et responsables des *policies* et du *politics*. Ce sont ces relations qu'il faut pouvoir travailler et faire évoluer pour sortir l'évaluation de son rôle mineur. J. Leca n'a cessé de relever le caractère central de cette articulation³³.

Ensuite, au vu de ce que les sciences sociales ont apporté depuis cinquante ans à l'analyse des processus de changement dans les institutions, ne faudrait-il pas dans un premier temps privilégier l'apprentissage des acteurs de l'action publique avant de prétendre modifier l'institutionnalisation ? Tous les observateurs, le disent et le répètent (et S. Jacob avec une netteté particulière) ce qui fait défaut en France ce sont les prémisses cognitives de l'action évaluative et l'absence de cette dimension dans les cultures professionnelles des acteurs publics³⁴. La réalisation de cette première suppose l'insertion d'une culture d'évaluation dans la formation ; elle peut aussi être accomplie (ou renforcée) par la découverte expérimentale des enjeux et problèmes de base de l'évaluation lors d'opérations internes mobilisatrices. Au sein de tous les dé-

33 J. Leca, « Sur le rôle de la connaissance dans la modernisation de l'État et du statut de l'évaluation », *Revue française d'administration publique*, 66, 1993, pp. 185-196. Également, P. Duran, « Connaissance et action publique, l'évaluation comme savoir pratique », *Penser l'action publique*, Paris, LGDJ, 1999, chap. IV, pp. 167-196.

Sur les illusions de la dépolitisation, D. Gaxie, P. Laborier, « Des obstacles à l'évaluation des actions publiques », in P. Favre et al., *Être gouverné*, Presses de Sciences Po, 2003, pp. 201-223.

34 Ce point est présent dès le rapport Viveret (*op. cit.* note 10) comme dans l'analyse internationale de Jacob et Varone (*op. cit.* note 4). Sur une formulation récente cf. V. Chanut, « L'évaluation affaire d'État ou questions d'organisation ? », *Politiques et management public*, 20 (4), 2002, pp. 1-32.

partements ministériels des opérations d'évaluation pourraient ainsi être menées à l'occasion des premiers suivis de la LOLF sur la base d'une méthodologie commune. Par exemple, tous les services de première ligne pourraient toutes s'efforcer de répondre à la question suivante : dans quelle mesure la nouvelle nomenclature budgétaire permet-elle de rendre comptes des actions menées ? Il s'agirait d'explicitier la réalité dont essayent de rendre compte les indicateurs choisis. Et de répondre aux questions de base : que mesure-t-on ? Comment cela est-il effectué ? Quelles significations sont ainsi produites ? Ce serait autant un moyen de mettre à l'épreuve les catégorisations et les indicateurs initialement retenus, qu'un exercice pédagogique d'apprentissage par l'épreuve d'une démarche évaluative. Parmi beaucoup d'autres, les travaux de K. Thelen³⁵ ont montré que les impacts exogènes (changement politique, réforme institutionnelle) produisent d'autant plus d'effets de changement qu'ils permettent l'actualisation et le renforcement de logiques déjà l'œuvre. C'est dans son modèle d'analyse la « sédimentation institutionnelle » qui permet de comprendre les possibilités et l'ampleur des changements observables. Dans le cas de l'évaluation, le choc exogène que constituerait la création d'une Haute autorité produirait d'autant plus d'effets qu'il aurait été précédé d'un travail d'apprentissage approfondi par la formation théorique et pratique.

La journée d'étude organisée par le CEVIPOF à l'initiative de la Direction scientifique de Sciences Po s'est efforcée de contribuer aux débats actuels dont nous venons de tracer les grandes lignes. Une première partie s'attache à l'analyse comparative d'un certain nombre de modèles d'organisation des activités d'évaluation, cela à partir du bilan d'une recherche comparative internationale et de la présentation de quelques cas (Québec, Suisse, France, Unesco) qui permettent de différencier les formes de commande politique et d'institutionnalisation. La deuxième partie apporte des éléments de réflexion sur la question des méthodes qui constitue avec le rapport au politique, l'autre débat récurrent à propos des démarches évaluatives. Nous verrons, une fois encore, à quel point les choix méthodologiques ne sont pas réductibles à des

35 K. Thelen, 2004, « Comment les institutions évoluent : perspectives de l'analyse comparative historique », *Année de la régulation*, 7 (16).

choix techniques, mais orientent de façon déterminante le contenu de la connaissance produite par l'évaluation (l'objet construit à partir de la commande et les dimensions qui prétendent en rendre compte). Si la méthode détermine la signification de la démarche, elle tient aussi une place centrale dans la légitimité de la l'opération menée, c'est-à-dire dans l'acceptabilité de ses informations et conclusions et finalement dans son impact sur les commanditaires et décideurs politiques. Enfin, dans une troisième partie nous présentons un texte original offrant une comparaison entre grands types d'évaluation.

PREMIÈRE PARTIE

BILAN COMPARÉ DE DÉMARCHES D'ÉVALUATION DES POLITIQUES PUBLIQUES

INTRODUCTION : L'INSTITUTIONNALISATION DE L'ÉVALUATION DANS UNE PERSPECTIVE COMPARÉE

Frédéric VARONE
UCL

F. Varone introduit la séance en présentant une première étude réalisée avec S. Jacob¹ sur les différentes pratiques d'institutionnalisation de l'évaluation dans 18 pays de l'OCDE. Il rappelle que l'évaluation est conçue à la fois comme un outil permettant de mesurer les effets propres d'une politique (enjeu méthodologique) et comme le moyen de porter un jugement de valeur sur ses mérites et ses effets (enjeu politique).

Il constate que la littérature actuelle sur le sujet regroupe des hypothèses parfois contradictoires. D'un côté, pour P. Duran, « *Le développement de l'évaluation est tributaire du type d'État dans lequel il se produit et de la nature des élites susceptibles de la porter*². » Il y aurait des États plus évaluateurs que d'autres. De l'autre, pour J.-C. Thoenig³, « *l'évaluation n'est pas par essence l'attribut d'un profil spécifique d'État ou de culture administrative* » mais un produit institutionnel qui évolue selon des demandes contingentes.

1 S. Jacob, F. Varone, « Institutionnalisation de l'évaluation et Nouvelle Gestion Publique : un état des lieux comparatif », *Revue internationale de politique comparée*, vol. 11, n° 12, janvier 2004.

2 P. Duran, « Les ambiguïtés politiques de l'évaluation », *Pouvoirs*, Paris, PUF, n° 67, 1993, p. 137.

3 J.-C. Thoenig, « L'évaluation, sources de connaissances applicables aux réformes de la gestion publique », *Revue française d'administration publique*, n° 92, 1999, p.684.

Il constate également que peu d'études finalement mesurent le degré d'institutionnalisation de l'évaluation dans différents pays, à l'exception notable de celle menée par J.-E. Furubo, R. Rist et R. Sandahl⁴. C'est pourquoi avec S. Jacob il propose plusieurs indicateurs pour tenter de mesurer ce degré d'institutionnalisation :

- *la pratique effective* (fréquence des appels politiques, part du budget consacrée à ce type d'activité...),
- *la présence ou non d'« incitants institutionnels »* (par exemple des clauses de lois rendant obligatoire l'évaluation d'un programme d'action),
- *les caractéristiques des structures organisationnelles* (instances isolées ou dispositif inter-organisationnel),
- *la présence ou non de « communautés épistémiques »* (présence ou non de clubs, de réseaux et de sociétés nationales d'évaluation, présence ou non de standards de qualité, de bulletin(s) d'information, revue(s) scientifique(s), réalisation d'une enquête bibliométrique).

Ces indicateurs sont ensuite synthétisés pour construire un indice théorique de mesure du degré d'institutionnalisation de l'évaluation. Le tableau suivant détaille la composition de cet indice, en la comparant à celle de l'indice construit par J.-E. Furubo, R. Rist, R. Sandahl.

⁴ J.-E. Furubo, R. Rist, R. Sandahl (eds), *International Atlas of Evaluation*, New Brunswick & Londres, Transaction Publishers, 2002.

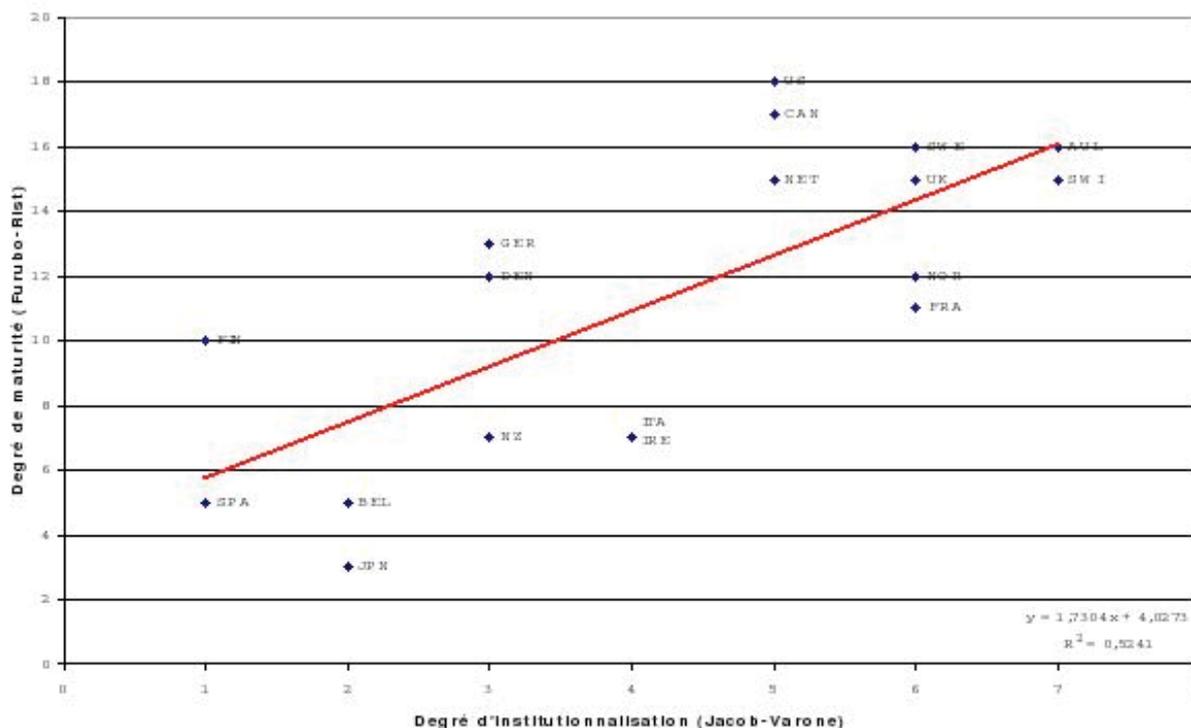
Tableau 1 : Composantes des indices théoriques des degrés de maturité et d'institutionnalisation de l'évaluation

	Institutionnalisation selon Jacob-Varone (Indice 0 à 9)	Maturité de la pratique selon List et Furubo (Indice 0 à 18)
INSTANCES D'ÉVALUATION		L'évaluation est répandue dans de nombreux secteurs de politiques publiques (de 0 à 2)
	Instance d'évaluation auprès de l'Exécutif (de 0 à 2)	Degré d'institutionnalisation – Gouvernement (de 0 à 2)
	Instance d'évaluation auprès du Parlement (de 0 à 2)	Degré d'institutionnalisation – Parlement (de 0 à 2)
	Pratiques de l'évaluation au sein de la Cour des Comptes (de 0 à 2)	Pratique de l'évaluation au sein de la Cour des comptes (de 0 à 2)
RÈGLES FORUM PRATIQUES		Soutien d'évaluateurs (nationaux) pluridisciplinaires (de 0 à 2)
	Société nationale d'évaluation (de 0 à 1)	Organisation professionnelle (de 0 à 2)
	Revue scientifique (de 0 à 1)	Discours national concernant l'évaluation (de 0 à 2)
	Standards de qualité (de 0 à 1)	Pluralisme des évaluations ou des évaluateurs au sein d'un secteur de politiques publiques (de 0 à 2)
		Prise en compte par l'évaluation des effets et de la mise en œuvre des politiques publiques (de 0 à 2)

Source : S. Jacob, F. Varone (2004)

Quel est alors le degré de développement de l'évaluation dans différents pays ? Pour répondre à cette question, F. Varone propose une approximation générale en croisant dans une analyse de régression deux indices théoriques présentés dans le tableau qui suit : le degré de maturité de la pratique d'évaluation (indice de J.-E. Furubo, R. Rist et R. Sandahl) et son degré d'institutionnalisation (indice de S. Jacob et F. Varone).

Figure 1 : Mesure empirique de la maturité et de l'institutionnalisation de l'évaluation dans dix-huit pays industrialisés



Source : S. Jacob, F. Varone, (2004)

L'idée générale est que plus on va vers la droite et le haut du graphique, plus l'évaluation serait institutionnalisée et pérennisée. On remarque alors que certains pays, comme les États-Unis, l'Australie, le Royaume-Uni, le Canada ou même la France pour qui l'indice est finalement flatteur, attestent d'un degré d'institutionnalisation et d'une maturité de la pratique évaluative élevés. Le tableau suivant synthétise les différents cas.

Tableau 2 : Répartition des pays en trois catégories

		Degré d'institutionnalisation	
DEGRÉ DE MATURITÉ		Élevé (si l'indice est supérieur à 4)	Bas (si l'indice est inférieur ou égal à 4)
	Élevé (si l'indice est supérieur ou égal à 11)	Angleterre, Australie, Canada, États-Unis, France, Norvège, Pays-Bas, Suède, Suisse	Allemagne, Danemark
	Bas (si l'indice est inférieur à 11)		Belgique, Espagne, Irlande, Japon, Nouvelle-Zélande, Italie, Finlande

Source : S. Jacob, F. Varone (2004)

À l'inverse, des pays comme la Belgique, la Nouvelle-Zélande, l'Espagne ou le Japon témoignent à la fois d'un degré de maturité et d'un degré d'institutionnalisation faibles. Il existe des pays de catégorie intermédiaire où le degré de pratique est élevé mais l'institutionnalisation faible (Allemagne, Danemark). En revanche il n'existe pas, d'après ces indices, de pays à faible maturité et forte institutionnalisation. L'analyse plus qualitative de S. Jacob sur le cas canadien confirme-t-elle cette présentation ?

LA PRATIQUE DE L'ÉVALUATION AU CANADA : LA RECHERCHE DE LA QUALITÉ

Steve JACOB
Université de Laval

La littérature présente souvent le Canada comme un pays novateur, « pionnier » même en matière d'évaluation. Celui-ci aurait des indices de maturité et d'institutionnalisation élevés. S. Jacob se situe en particulier au niveau fédéral. À ce niveau en effet, il remarque qu'il existe une recherche constante de l'utilité et de la qualité des programmes mis en place. Mais si l'on insiste sur le *processus* même d'institutionnalisation, on peut nuancer cette image. L'auteur nous invite alors à remonter aux origines du processus.

Même s'il y a débat sur ces origines (remontent-elles au XIX^e siècle ? à 1931 avec la loi sur le revenu consolidé et la vérification ?), on fait naître en général l'institutionnalisation de l'évaluation au Canada avec la circulaire du Conseil du Trésor¹ de 1977 intitulée « *Évaluation de programmes effectués par les ministères et les organismes* ». Quels éléments ont contribué à l'apparition de cette circulaire ? Cette question est centrale car sa réponse montre que l'évaluation est d'abord la réponse aux problèmes d'une époque.

Il y a tout d'abord des acteurs porteurs comme Maxwell Anderson qui est vérificateur général du Canada en 1960, c'est-à-dire agent du Parlement chargé de contrôler les dépenses gouvernementales. M. Anderson étend la compétence du vérificateur général en introduisant fréquemment ses propres com-

¹ Pour fixer les idées, le Conseil du Trésor au Canada serait l'équivalent en France d'un ministère regroupant le ministère de la Fonction publique et le ministère des Finances.

mentaires dans ses rapports d'activité. Une telle pratique a plusieurs effets favorisant l'institutionnalisation de l'évaluation :

- elle attire l'attention des médias et contribue à populariser son travail,
- elle oriente l'évaluation sur la voie de la recherche de la performance,
- elle permet la mise sur agenda, ou plus précisément la mise en place d'une commission royale d'enquête sur le sujet qui préconise l'instauration d'un système de reddition des comptes publics. Pour les membres de cette commission en effet, la validation des dépenses gouvernementales n'est pas seulement une question de légalité. Il s'agit de promouvoir ce qu'on a pu appeler les « trois E » : efficacité, efficience et économie. Cette première tentative échoue, mais elle favorise le recrutement de nouveaux cadres formés au management.

Le contrôleur général lui aussi publie dès 1976 un rapport dans lequel il s'inquiète de voir le Parlement perdre le contrôle des dépenses publiques. Il propose la mise en place d'indicateurs de performance. Une commission d'enquête royale est créée mais elle ne remet pas son rapport puisque la circulaire mentionnée précédemment est publiée entre temps.

Cette circulaire prévoit plusieurs finalités pour l'évaluation : faciliter l'affectation des ressources, améliorer les « trois E », conduire à l'amélioration de programmes (rendement des administrations, atteinte des objectifs...) et améliorer l'imputabilité. L'évaluation devient alors un instrument d'aide à la décision au service des gestionnaires publics. Elle est réalisée tous les 5 ans. De même les rôles sont bien délimités. Les hauts fonctionnaires ont la liberté de développer et d'organiser les structures adéquates à leur fonction. Il n'y a pas d'obligation institutionnelle en la matière. Ce sont eux qui impulsent le processus d'évaluation en créant des services internes d'évaluation. Le vérificateur général quant à lui s'assure de la bonne conduite de ces évaluations internes et publie des guides sur l'évaluation. Le contrôleur général enfin publie fréquemment des rapports stigmatisant ceux qui ne procèdent pas aux évaluations nécessaires.

Mais il manque à cette circulaire un aspect incitatif. Le Conseil du Trésor demande alors les planifications annuelles des évaluations et les résumés des évaluations réalisées. Il conditionne l'obtention des fonds budgétaires au res-

pect du cycle quinquennal des évaluations. Il convient de rappeler ici le contexte de fonds, à savoir la crise économique des années 1980 et la nécessité de réduire les dépenses publiques. Ce contexte a une incidence sur les méthodes employées : « value for money », calcul coûts/avantages, calcul coûts/efficacité...

Pourtant dans les années 1980, une étude de la Commission sur les coupes budgétaires menées à cette époque conclut que ces coupes ont été réalisées de façon routinière, voire parfois aléatoire, sans critère d'ensemble et donc que l'évaluation fut inutile. C'est pourquoi une réforme de ce système est lancée au milieu des années 1990 qui focalise sur les notions de rendement et de productivité. Un nouveau rôle est assigné à l'évaluation : celle-ci doit proposer une alternative pour ne pas enfermer la décision publique dans une simple logique de financement comptable. Elle touche alors à l'essence même de l'action publique.

Actuellement l'évaluation au Canada désigne à la fois la mesure des rendements et la production de renseignements précis et détaillés pour aider les gestionnaires dans leurs décisions et influencer la population. L'évaluation est considérée comme un moteur du changement. En revanche elle ne participe pas à l'amélioration de la transparence administrative car la plupart des rapports ne sont pas rendus publics. L'explication en est simple : il est admis que les gestionnaires sont peu enclins à donner des informations menant à la critique de leur métier. Le politique a dès lors privilégié les instruments de changement interne.

Toutefois un centre d'excellence en évaluation auprès du secrétariat du Conseil du Trésor fut créé pour contrôler le développement de ces évaluations internes. Ce centre consacre à cette tâche 1,5 millions de dollars alors que, pour comparer, le budget total consacré aux évaluations internes s'élève à 4 millions. En outre il existe depuis 1981 une Société canadienne d'évaluation qui publie une revue scientifique en anglais et en français et qui a adopté récemment un code d'éthique des évaluateurs.

Concernant l'usage de ces évaluations, on constate que la plupart des rapports sont lus avant la publication plutôt qu'après. Une étude du *Globe and Mail* a

été réalisée pour mesurer l'utilisation réelle des évaluations par les évalués. D'après cette étude, 3 % des évaluations seulement ne sont pas utilisées. 4,5 % des évaluations mènent à une suppression du programme évalué, 15 % à une réforme radicale, 40 % conduisent à des modifications opérationnelles, 10 % à la confirmation du programme en cours et 28 % participeraient d'une dynamique d'apprentissage intra-organisationnel.

Pour S. Jacob, le défi du Canada des prochaines années est de bien sensibiliser les gestionnaires à l'utilité de ces évaluations et de donner aux évaluateurs une identité propre. Tout le monde peut se prétendre évaluateur aujourd'hui au Canada. La Société canadienne d'évaluation réfléchit alors à un processus de certification des évaluateurs pouvant mener à ce qu'on pourrait appeler un « ordre des évaluateurs ». S. Jacob conclut en précisant que malgré les critiques incessantes adressées aux évaluations, il existe bien une fonction évaluative, qui elle, n'est pas remise en question et qui évolue pour le moment de façon incrémentale au fil des débats la concernant.

LES PRATIQUES D'ÉVALUATION DANS LES ORGANISATIONS INTERNATIONALES

Florent LOISEAU
UNESCO

Pour faire le lien avec l'intervention précédente, F. Loiseau rappelle que le vérificateur général du Canada a joué un grand rôle à l'UNESCO où il a contribué à la mise en œuvre du service d'évaluation interne et où il remplit actuellement la fonction de vérificateur général externe, donc d'évaluateur des évaluateurs internes. Pour bien comprendre le positionnement de cette organisation au sein des organisations internationales, l'auteur introduit quelques distinctions élémentaires.

D'une part il faut bien distinguer les organisations non gouvernementales (ONG) des organisations intergouvernementales. D'autre part il faut différencier parmi ces dernières les organisations régionales (comme l'Union européenne par exemple) des organisations multilatérales (comme l'ONU) Enfin, au sein de celles-ci, on distingue généralement les organisations issues du système de Bretton Woods de celles qui relèvent du système des Nations unies comme les Fonds (UNICEF par exemple), les Programmes (comme le PNUD) ou les institutions spécialisées (comme l'UNESCO). Ces distinctions sont importantes car on ne manie pas les mêmes programmes d'un cas à l'autre. À l'UNICEF par exemple, on mesure des services rendus à des pays ou des agences. À l'UNESCO, on produit des idées qui ne sont pas toujours en lien direct avec les pays en développement concernés. De même on met moins l'accent sur la mise en œuvre effective que sur le service rendu aux États membres. Les évaluations ne sont donc pas conçues de la même manière d'un cas à l'autre.

Une fois ces précisions apportées, F. Loiseau dresse un constat général de l'évolution de cette organisation. Tout d'abord il observe lui aussi au niveau international que le travail de l'UNESCO évolue, d'un souci de conformité à son acte constitutif vers une évaluation de la qualité du service rendu. Il observe également une remise en cause de certains usages de l'argent public. Les exigences d'efficience, d'efficacité, de transparence, les impératifs du « New Public Management », le « value for money » s'appliquent bien à l'évolution de l'activité de l'UNESCO. Il observe également la mise en place de nouveaux outils de gestion au sein de cette organisation :

- *la programmation par les résultats* l'emporte sur la production de « ce qu'on sait faire ». Aujourd'hui par exemple, on insiste davantage en interne sur le nombre de sites devant être inscrits sur l'inventaire et devant être réhabilités d'ici 2 ans,
- en externe, des *objectifs chiffrés* doivent être atteints,
- *des plans nationaux ou régionaux* (surtout en éducation) sont mis en œuvre, des plans de réduction de la pauvreté également (surtout par la Banque mondiale). On insiste plus sur les résultats et le degré de communication entre les receveurs et les agences internationales.

Plus précisément, l'auteur met en évidence trois grandes tendances au sein de l'UNESCO.

1. LE RECOURS CROISSANT À DES ÉVALUATIONS EXTERNES

Son service, auparavant, était couplé avec le service de planification stratégique. Puis il fut rattaché à un service d'audit pour former le service de contrôle interne. L'UNESCO cherche un regard extérieur sur sa mise en œuvre, mais un regard tout de même averti. On assiste alors à une « externalisation en interne ».

En parallèle, le nombre des évaluateurs totalement externes augmente. Ces évaluateurs appartiennent aux États contribuant le plus au budget de l'organisation et sont financés par des ressources extrabudgétaires, extérieures à l'UNESCO. La part de ce financement dans le budget que les États membres

consacrent à l'UNESCO ne cesse d'ailleurs d'augmenter. Ces États peuvent dès lors mieux imposer leur vision des évaluations.

Il existe une croyance partagée selon laquelle l'extériorité est synonyme d'objectivité et limite les conflits d'intérêts. La notion même d'objectivité n'est pas remise en cause. Enfin une autre raison explique le développement de ces évaluations externes : la pluralité des offreurs, qui est un avantage pour le demandeur qu'est l'UNESCO. Cette externalisation de l'évaluation se traduit par :

- *une réorganisation des services internes.* Le service de F. Loiseau n'exécute plus que des évaluations de programmes strictement internes ou des évaluations nécessitant une expérience particulière sur un dossier précis, donc une part résiduelle de l'ensemble des évaluations menées. Son travail devient un travail de second rang qui consiste à participer au choix des programmes à évaluer, à piloter l'implication des responsables (de l'évaluation ou ceux qui sont évalués), à rédiger des termes de référence, à participer au choix des évaluateurs externes et à suivre les rapports d'évaluation (contrôle méthodologique, impact),
- *une évolution des méthodes.* Les évaluateurs externes, par leur expérience, leurs connaissances, leur réseau peuvent plus facilement faire des comparaisons générales. De plus les évaluateurs coûtent cher (on ne peut pas les envoyer partout dans le monde). On voit alors se développer des méthodes d'évaluation secondaires (revues de littérature, interviews par téléphone, questionnaire, téléconférences...) et une diminution des observations de terrain, du face-à-face, et des connaissances des mécanismes internes aux organisations.

2. UNE PROFESSIONNALISATION DE L'ÉVALUATION

L'auteur préfère le terme de profession à celui de communauté épistémique. Dans un contexte d'augmentation de la demande d'évaluation, il existe un espace intermédiaire pour ces corps d'experts qui veulent se constituer en profession.

Ce phénomène se voit à travers le développement de la formation professionnelle (cycles de formation internationale dans les universités ou à Sciences Po

par exemple), le développement de la formation continue (exemple de l'« Evaluation Institute » à Washington), le développement de sociétés (nationales le plus souvent) d'évaluation, la standardisation croissante des pratiques évaluatives (produites par l'OCDE ou plus récemment par le groupe évaluation des Nations unies) et le raffinement méthodologique et théorique (production constante de nouveaux paradigmes, valorisation des « theory based approach », développement des méta-évaluations, développement des « peer review », évaluation par les pairs).

3. UNE AMBIGUÏTÉ CROISSANTE DES RAPPORTS ENTRE LES METTEURS EN ŒUVRE (LES SPÉCIALISTES DE PROGRAMME À L'UNESCO) ET LES ÉVALUATEURS

L'évaluation devient incontournable. Elle est cyclique, elle conditionne l'obtention de fonds... La pression est donc plus forte sur la mise en œuvre. Mais en parallèle on assiste à une raréfaction des ressources consacrées à la mise en œuvre ! Et quand bien même les ressources existent, on observe une compétition croissante pour leur obtention. On a pu le voir dans le cas de l'éducation primaire après le récent tsunami. Par conséquent les metteurs en œuvre prétendent qu'il vaut mieux mettre l'argent dans l'éducation que dans l'évaluation. Et ils se forment en parallèle à l'évaluation, surtout interne, pour pouvoir mieux se préparer aux évaluations de leur propre travail par des évaluateurs externes.

L'auteur conclut alors à un rapprochement croissant des fonctions d'évaluation classique et d'audit. Les formations et les méthodes sont souvent conjointes. Les objets sont parfois hybrides : évaluation récurrente de programmes, audits organisationnels, audits de performances... Il conclut également en mettant en évidence l'incompréhension croissante qui s'installe entre les évaluateurs, qui ont du mal à s'émanciper du contrôle budgétaire et qui ont peu de connaissances en sciences sociales, et les membres de son service.

F. Loiseau met enfin l'accent sur les dérives possibles d'une telle situation que sont l'excès de formalisme et la croissance de la bureaucratisation. Il constate d'ailleurs que la situation de l'évaluation est contrastée : toujours plus fortement institutionnalisée, elle a pourtant du mal à jouer son rôle dans la réforme gestionnaire car elle a un accès limité aux sources internes et peu de retour de la part des États membres. Il propose alors deux pistes pour améliorer la situation :

- *insérer davantage l'évaluation dans un processus de programmation.* Le service de F. Loiseau a en effet du mal à revenir sur la programmation après l'évaluation car à ce moment-là, ce service serait à la fois juge et partie. Et en même temps, il reconnaît que l'évaluation n'a de sens que si ce retour est réalisé,
- *plus impliquer les bénéficiaires ultimes de l'évaluation dans le processus évaluatif.* Cette action est toutefois rendue difficile par le statut même des évaluateurs externes (peu disponibles, manque de temps).

DÉBAT AVEC L'ASSEMBLÉE

F. Varone introduit les débats en rappelant qu'une des conclusions intéressantes de l'étude de S. Jacob est de montrer que, même si elle ne permet pas d'améliorer la transparence de l'administration, l'évaluation permet dans près de 20 % des cas un changement radical de politique...

Quels sont les acteurs qui constituent les dites « communautés épistémiques » dans les différents pays ? Ne s'agit-il que d'évaluateurs ? Quel rôle jouent les commanditaires ? Ne minore-t-on pas le rôle de grands pays du Nord ou de certains cabinets d'experts ?

La réponse de F. Varone est nuancée. Elle dépend en fait du pays considéré. Il remarque qu'en Suisse par exemple, on rassemble l'ensemble des parties prenantes avant l'évaluation.

S. Jacob insiste sur l'aspect politique de l'évaluation. Au Canada, celle-ci a été utilisée uniquement par le gouvernement. Le Parlement, dont les pouvoirs diminuent, ne voit pas encore en l'évaluation un moyen de recouvrer ces derniers.

Par ailleurs, il a mené, à l'occasion du vingtième anniversaire de la *Revue canadienne d'évaluation et de programme*, une analyse bibliométrique des 400 articles qu'elle avait déjà publiés. Cette analyse montre que les auteurs sont d'abord des Américains. Puis on observe la publication de quelques auteurs canadiens qui s'adaptent à ce nouveau lexique. Les auteurs européens apparaissent seulement à partir des années 1990 quand ce thème se développe en Europe. 60 % des articles sont rédigés par des universitaires. Concernant les sujets abordés, aucun article n'évoque la coopération et le développement ou les affaires étrangères. La majorité d'entre eux concernent l'éducation ou les politiques sociales.

Pour F. Loiseau, les cabinets d'experts jouent un rôle mineur dans le cas de l'UNESCO. Il existe une forte incitation à évaluer l'utilisation du budget à terme mais il est rarement précisé qui doit faire cette évaluation et sur quelles bases. C'est là que le service de F. Loiseau se trouve dans une situation difficile. Ce service doit contrôler les évaluateurs externes pour garantir la qualité de leur travail. Mais ce n'est pas toujours le cas, les acteurs chargés de la mise en œuvre contactant directement les organismes extérieurs chargés de l'évaluation.

S. Le Bouler (Commissariat général du Plan) témoigne de son malaise devant certaines labellisations comme les évaluations « externes/internes » ou les « communautés épistémiques », dans la mesure où l'on a ouvert au maximum les marchés publics et où on crée par là des marchés captifs (essentiellement en France le marché des contrats de plan État-régions et le marché des fonds structurels). Il se réjouit qu'on ait refusé cela en France pour éviter un renfermement du secteur de l'évaluation sur le milieu des administrateurs.

Quelle est l'influence des évaluations lancées par l'Union européenne sur le développement de l'évaluation ? Observe-t-on par exemple un lissage des pratiques dans les différents États ?

F. Varone se fonde sur la liste des facteurs d'institutionnalisation de l'évaluation isolés dans son enquête pour répondre à cette question. Ces facteurs sont :

la Nouvelle gestion publique (est-elle un frein à l'évaluation ?),
le poids des déficits budgétaires (évaluer pour mieux couper ?),
les contraintes extérieures, entre obligation réglementaire et initiative propre,
la présence d'un régime majoritaire (l'évaluation comme levier de l'opposition ?),
l'indépendance des organes de contrôle (l'évaluation comme nouvel outil ?),
la présence d'une structure fédéraliste (un laboratoire pour l'évaluation ?).

Il remarque qu'en effet dans les cas italien et espagnol, rien ne peut se penser sans l'aide de l'Union européenne et donc sans l'effet d'une contrainte extérieure. Dans la plupart des cas, l'obtention des fonds structurels est conditionnée. On observe dans ces pays « récipients d'air » que les autorités, qui vivent d'abord l'évaluation comme une contrainte, se l'approprient ensuite et la diffusent auprès de leurs partenaires.

QUE FAIRE DU CAS DE LA FRANCE ?

La plupart des participants sont surpris de la place faite à la France dans la typologie présentée par F. Varone. J.-P. Nioche (professeur à HEC) affirme par expérience que la France n'a jamais été dans le haut du tableau ! P. Lascoumes lance le débat sous la forme d'une boutade : la France n'est-elle pas l'exemple même d'un pays dans lequel l'évaluation serait fortement institutionnalisée sans faire preuve pour autant d'une grande « maturité » ?

S. Le Bouler trouverait éclairant qu'on puisse retirer de l'ensemble des évaluations prises en compte dans la présentation de M. Varone les évaluations des politiques contractuelles et les évaluations relevant des fonds structurels pour pouvoir mesurer le véritable degré d'institutionnalisation des évaluations. Cela permettrait notamment de situer la France dans les cas les plus faibles.

F. Varone rappelle que ce sont les Français qui ont placé la France où elle est. Concernant l'impact des fonds structurels et leur lien avec le développement d'une politique d'évaluation, il précise, son enquête à l'appui, que ce lien ne fonctionne pas toujours, qu'il n'est pas linéaire. Dès lors il comprend la sug-

gestion de S. Leboulet mais envisage également que ce retrait n'ait pas autant d'effet que cela.

ET DU CANADA ?

J.-P. Nioche remarque concernant le Canada, que S. Jacob a essentiellement parlé du niveau fédéral. Or il demeure frappé, pour avoir fréquenté les Sociétés canadienne et québécoise d'évaluation, par le particularisme québécois et surtout la prégnance des idées du Commissariat général au Plan dans ces associations. Quelle est dès lors la perception au niveau fédéral du particularisme québécois ?

S. Jacob répond en précisant qu'il n'existe pas uniquement un style québécois d'évaluation, mais bien autant de styles que de provinces. Il n'observe pas en outre de liens entre le niveau fédéral et le niveau des provinces. Peu de transferts de connaissance sont réalisés et il existe une barrière nette entre Ottawa et Québec. Il est frappant de constater par exemple que la Société québécoise d'évaluation, certes très active, n'a pas encore d'ancrage institutionnel clair. Au Québec les évaluateurs doivent à chaque fois se battre pour voir l'évaluation inscrite sur l'agenda politique. Par exemple la première version de la dernière loi de modernisation administrative ne mentionnait pas le rôle de l'évaluation et il a fallu que les évaluateurs luttent pour cela.

Enfin, ne donne-t-on pas un pouvoir immense aux évaluateurs ? Comment modérer alors ce pouvoir, notamment entre la phase de production et la phase de diffusion du rapport d'évaluation ?

F. Loiseau répond du point de vue de l'UNESCO. Dans cette organisation, les résultats des rapports ne sont pas tous diffusés et surtout ils sont rarement diffusés tels quels aux États membres. C'est son service interne d'évaluation qui rédige et publie la synthèse des rapports qui peuvent être publiés. En outre, il rappelle que ce service a une influence sur les évaluateurs externes dans la mesure où il suit étape par étape leur travail. Il existe un premier contrôle avant même la diffusion. S'ouvre ensuite une discussion en interne. Son service peut trouver qu'un rapport est moyen, mais parce qu'un ensemble

d'acteurs ont intérêt à le trouver bon, il peut le présenter comme tel. À cela s'ajoute le jeu des États membres qui peuvent accepter, reconnaître, valoriser tel ou tel rapport. Si tel État trouve que le programme solidement évalué comme un mauvais programme est utile, il n'est pas tenu de l'arrêter et le service de F. Loiseau ne peut rien faire.

S. Jacob ajoute qu'il est fort possible de surestimer le pouvoir des évaluateurs. Ceux-ci ont tout de même un cahier des charges qui encadrent leur travail. Il y a également un filtre politique quant à la faisabilité des recommandations. Ce filtre est souvent indépendant des exigences de qualité. S. Jacob cite comme exemple un passage d'un de ses entretiens avec un président de région socialiste en France qui déclare pour justifier la non prise en compte d'un rapport d'évaluation : « J'ai un problème avec ce rapport car il ne me convient pas. J'ai l'impression qu'ils n'ont pas compris mes questions ».

Pour F. Loiseau, il existe aussi un filtre cognitif. L'acteur peut être honnête et ne pas comprendre pourquoi les évaluateurs se focalisent sur tel point.

F. Varone approuve. En effet il ne faut pas surestimer le pouvoir des évaluateurs. Il rappelle que, dans les pays de l'OCDE, seulement 1 % en moyenne du budget prévu pour le lancement d'une politique est consacré à l'évaluation. Il remarque que beaucoup de pays se motivent mais que les effets sont souvent faibles. Pourtant, il est frappant de constater qu'il existe toujours des enjeux très forts de réappropriation politique de ces évaluations.

Pour J.-P. Nioche, le pouvoir des évaluateurs est même un fantasme sans fondement. Pour lui, la seule différence que l'on puisse faire, c'est si l'évaluation est publique ou non. Si le commanditaire a décidé dès le départ de rendre l'évaluation publique, certes il doit justifier par le discours le moindre de ses actes et s'en trouve bien souvent embarrassé. Mais si ce n'est pas le cas, l'évaluateur a peu de recours.

DÉVELOPPEMENT ET BILAN DES PRATIQUES D'ÉVALUATION AU SEIN DU CANTON DE GENÈVE

Jean-Daniel DELLEY
Université de Genève

J.-D. Delley attire l'attention de l'assemblée sur le « micro cas » que constitue le canton de Genève, tant du point de vue de l'échelle concernée (territoire restreint, population modeste) que du type d'institutions en jeu (Genève est une République qui n'est qu'un canton).

Pour lui, ce micro cas mérite notre attention pour au moins deux originalités. La première est que le canton de Genève a institutionnalisé l'évaluation de ses lois en créant en 1985 un organe permanent et externe à l'administration : la CEPP (la Commission externe d'évaluation des politiques publiques) dont l'auteur fut le président pendant 8 ans. La deuxième est que cette CEPP dispose d'un droit d'auto saisine et qu'en théorie elle ne dépend pas des autorités politico-administratives. Il nous propose alors une présentation en deux temps. Après avoir fait le portrait institutionnel de la CEPP, il nous en explique le fonctionnement.

La CEPP est composée de 16 membres désignés par le gouvernement pour un mandat de 8 ans, non renouvelable. Elle reçoit ses mandats du gouvernement ou de deux commissions spécialisées du Parlement. Ses membres sont choisis en dehors de l'administration et de toute autorité politique du canton au sens large. Par exemple l'administrateur d'un hôpital cantonal universitaire ne peut pas être membre de la CEPP. Ils sont choisis pour leurs « compétences particulières en matière de gestion économique et politique ». Dès la création de la CEPP, les instigateurs ont instauré un mélange dans la composition de cette commission qui s'est avéré très pratique et que la tradition a maintenu :

7 ou 8 membres proviennent des milieux économiques (patrons, directeurs de banque, indépendants...),

un autre bloc est constitué par des universitaires supposés s'assurer de la rigueur des exercices menés,

le dernier bloc se compose d'anciens députés ou magistrats¹, voire de journalistes.

Il s'agit donc d'une « commission de milice », c'est-à-dire d'une commission censée utiliser toutes les forces de la société civile, même si ces forces ne sont pas des professionnels de l'évaluation.

La CEPP a un budget propre pour financer l'activité de ses membres et de son secrétariat exécutif d'évaluateurs (regroupant 1,8 postes...), ses investigations et ses publications. Son pouvoir d'investigation est presque sans limite. D'un point de vue politique, la CEPP est seule maîtresse et responsable du produit final. Ses rapports, qui présentent la démarche utilisée, les résultats de l'évaluation et différentes recommandations, sont rendus publics à partir de 1985, du fait de l'accord tacite du ministère de cette époque. Pour J.-D. Delley d'ailleurs, il ne peut exister de véritable évaluation tant que celle-ci n'est pas rendue publique².

Comment fonctionne la CEPP ? J.-D. Delley rappelle d'abord que ce fonctionnement fut instauré par les évaluateurs eux-mêmes, la CEPP étant une création nouvelle. Concrètement, 3 ou 4 membres se saisissent d'un dossier jusqu'à son terme, avec l'appui du secrétariat exécutif. Ce groupe doit tout d'abord produire un premier rapport de quelques pages pour convaincre la commission qu'un tel travail est justifié. Puis il procède à une étude de faisabilité, elle aussi validée par la commission. Une fois l'évaluation lancée, le groupe d'évaluateurs doit présenter tous les mois à la commission un rapport d'activité auquel celle-ci est amenée à réagir. Cette contrainte – J.-D. Delley n'hésite pas à dire que « le groupe est tenu en laisse par la commission » – est aussi un réconfort, chaque évaluateur se sentant épaulé par toute la commis-

¹ J.-D. Delley désigne par là les hommes politiques à la tête de communes, et non des juges...

² Tous les rapports de la CEPP sont d'ailleurs disponibles à l'adresse suivante : www.geneve.ch/cepp.

sion dans son travail. Le rapport final d'évaluation est lui aussi présenté à la commission, qui doit l'approuver à l'unanimité pour que celui-ci soit publié.

J.-D. Delley insiste sur le fait que, contrairement à certains instituts universitaires, la CEPP a toujours été soucieuse de l'utilité de son travail. C'est pourquoi aucune évaluation ne doit durer plus d'un an, ce qui permet en outre d'éviter des dépenses abusives. La CEPP ne peut alors choisir qu'un nombre limité de questions et n'évaluer que 2 ou 3 points saillants pour les autorités ou pour la CEPP elle-même. Pour J.-D. Delley, prétendre évaluer toute une politique publique est souvent du bluff et ne revient la plupart du temps qu'à faire le bilan de ce qui existe...

Lors de ses évaluations, la CEPP rencontre rarement des politiques publiques dont les objectifs sont explicites, à l'exception notoire de la politique des transports urbains qui prévoyait qu'aucun transport urbain ne devait dépasser les 18 km/h. Même dans ce cas pourtant, les membres de la CEPP se sont rendu compte que saisir le degré d'atteinte d'un objectif est un non-sens et que la politique évaluée ne se résume pas à ses objectifs. La CEPP se livre alors à un diagnostic de situation dont le but est de montrer qui bénéficie (ou non) de la politique mise en œuvre. Ce diagnostic a souvent pour effet premier d'informer l'administration sur sa propre action. J.-D. Delley prend l'exemple des logements publics, dont l'administration ignorait la cherté, cherté qui dépassait parfois celle des logements privés !

Cette démarche consiste pour lui à « tendre un miroir » à l'administration sur son activité. L'évaluation n'est ni le moment ni le lieu d'une révolution. Pour J.-D. Delley, les membres de la CEPP ont dû apprendre avec le temps à limiter l'ambition de leurs recommandations, quitte à énoncer parfois de véritables truismes et à se limiter à une description détaillée du terrain. Pour être un bon évaluateur, il faut au minimum savoir faire cela. Mais il ne faut pas aller plus loin, sinon l'évaluateur entre dans une logique politique qui n'est pas la sienne. La CEPP est d'ailleurs de toutes tendances politiques. C'est pourquoi il faut que l'avis final de la commission soit unanime, pour éviter des recommandations politiques. C'est finalement ce système de milice à l'origine de la composition de la commission qui permet pour J.-D. Delley de concilier la

dimension politique et la dimension méthodologique de l'évaluation : certes les membres sont encadrés par des universitaires, mais ils ont aussi des expériences politiques et ils sont tous des citoyens évoluant dans le milieu qu'ils évaluent.

Ce positionnement apolitique si difficile à tenir illustre bien l'ambiguïté générale du métier d'évaluateur pour J.-D. Delley. L'évaluateur doit tenir un équilibre délicat entre deux positions peu compatibles. S'il est vraiment un évaluateur externe, il doit savoir être indépendant de l'administration. Mais le risque pour lui est alors de ne pas remplir sa deuxième fonction qui est celle de conseil et qui implique des rapports étroits avec les autorités. Il est alors constamment sur le fil du rasoir et il doit apprendre à jongler entre le besoin d'autonomie nécessaire à toute action crédible d'une part, et la proximité nécessaire à la faisabilité de ses recommandations d'autre part. Pour autant, J.-D. Delley conclut par une pensée personnelle qui peut paraître surprenante au vu de ce qui vient d'être énoncé : pour lui l'évaluation reste avant tout un acte politique.

L'ÉVALUATION DES POLITIQUES PUBLIQUES AU COMMISSARIAT GÉNÉRAL AU PLAN. RÉCIT D'UNE EXPÉRIENCE

Stéphane LE BOULER
Commissariat général du Plan

Après s'être rapidement défini comme un « intermittent de l'évaluation », S. Le Bouler nous invite à écouter l'histoire du constat « clinique » de l'évaluation telle qu'elle a pu être vécue au niveau interministériel. La fonction d'évaluation interministérielle est instaurée dès 1990 et est rattachée au Commissariat général du Plan. Elle se termine à l'automne 2003. Il est donc possible de tirer les leçons des différents dispositifs interministériels mis en place entre ces deux dates.

Le premier dispositif mis en place en 1990 avec l'instauration du Comité interministériel de l'évaluation et du Conseil scientifique de l'évaluation (dispositif CIME/CSE) a fait l'objet de nombreuses critiques : lourdeur administrative des procédures, déperdition des projets, délais d'évaluation trop longs, caractère périphérique des sujets, difficultés d'instruction, médiocre gestion des suites de ces évaluations et faible implication des commanditaires.

Par conséquent le décret du 18 novembre 1998 révisé cette procédure d'évaluation et prévoit de réduire les délais, de simplifier le processus d'instruction et d'élargir la saisine du Comité national d'évaluation (CNE, qui remplace le dispositif CIME/CSE). Toutefois pour S. Le Bouler, il ne s'agit là que d'une simple révision institutionnelle, ce qui explique qu'on retrouve parmi les critiques adressées au nouveau dispositif au début des années 2000 bien des critiques énoncées ci-dessus.

Le dispositif de 1998 a lui aussi été réformé. On s'assure notamment de la publicité des rapports des évaluations menées à leur terme. Plusieurs transformations sont également apportées à la procédure de programmation sous l'égide du CNE :

la diversification du droit de saisine du CNE. S. Le Bouler observe qu'il s'est agi là simplement d'éviter que ce ne soit uniquement les ministères de l'Emploi, du Travail, de la Santé et de l'Éducation nationale qui commandent des évaluations mais bien tous les ministères. Il ne s'est jamais agi d'accorder le droit de saisine au Parlement, ce qui fut ensuite reproché au dispositif,

la « montée en gamme¹ » du nombre des évaluations (voir le détail des thèmes dans l'encadré ci-dessous). Mais l'auteur considère que ces évaluations concernaient encore des politiques ou des tranches d'action publique trop vastes et mal identifiées a priori pour pouvoir déboucher sur des évaluations sérieuses dans un temps si court et avec un personnel restreint.

Thème des évaluations au programme du CNE

Année 1999 :

- la politique de préservation de la ressource en eau,
- les dispositifs d'emplois-jeunes, de Contrat emploi solidarité et de Contrat emploi consolidé,
- la politique de logement social dans les Départements d'Outre Mer (cette évaluation est la seule qui ne fut pas menée à son terme),
- les mesures d'aide au secteur d'emploi non-marchand,
- la politique de lutte contre le SIDA.

Année 2000 :

- la politique de sécurité routière (qui a débouché sur la mise en place des limiteurs de vitesse),
- la formation professionnelle et continue des agents de l'État,

¹ Les termes sont de S. Leboulet.

- les politiques de développement rural.

Année 2001 :

- la politique d'aide aux très petites entreprises,
- le service public des déchets ménagers,
- les fonds structurels et les politiques régionales,
- la politique de contractualisation avec les universités,
- le recours à des opérateurs externes en matière de politique d'emploi,
- la politique de transport combiné rail/route,
- les politiques d'amélioration de l'accès à la prévention et aux soins.

la question des suites. En la matière, S. Le Bouler remarque que le dispositif n'était guère différent de celui de 1998 qui prévoyait une phase de gestion administrative des évaluations entre la remise du rapport et son utilisation effective. Toutefois la pratique a changé puisque les évaluateurs, conscients des lenteurs administratives, communiquaient directement aux commanditaires les conclusions de l'évaluation dans le but de permettre ce que l'auteur appelle « la rencontre improbable » entre les besoins des décideurs politiques et la production d'information des évaluateurs. Il remarque également que l'alternance politique de 2002 a été profitable, le nouveau gouvernement utilisant volontiers les évaluations commandées par le gouvernement précédent pour légitimer son action.

Pourquoi dès lors le CNE, malgré ses multiples réformes, fut toujours critiqué ? S. Le Bouler émet l'hypothèse que cette situation provienne d'une incompréhension persistante entre les Parlementaires et le Commissariat général du Plan sur le fonctionnement même du dispositif. Au cours des évaluations menées par les différentes instances et coordonnées par le CNE, il était fréquent de voir ces dernières recourir à des évaluateurs externes sur des points particuliers même si l'évaluation finale leur appartenait toujours. Cette division des tâches n'était pas claire pour certains élus : elle était interprétée comme une organisation coûteuse et finalement plus compliquée qu'une véritable évaluation externe, sur la base d'indicateurs quantitatifs notamment. C'est pourquoi, par la suite, cette externalisation partielle de l'évaluation fut

plus codifiée, des budgets, certes peu contraignants, étant définis pour chaque évaluation ponctuelle extérieure.

Les délais constituent un autre élément d'explication. Critique adressée de façon récurrente au dispositif, ces délais prenaient plusieurs formes. La programmation annuelle du ministre nécessaire au lancement de l'évaluation par exemple prend nécessairement du temps. Il en va de même quand l'évaluation suppose le recueil de l'avis de différentes personnes évaluées.

Le caractère bicéphale du dispositif ne joue pas non plus en sa faveur avec d'un côté le CNE qui se dote d'un programme d'évaluation et définit la composition de l'instance d'évaluation chargée de la conduire, et de l'autre le Commissariat général du Plan qui assure le secrétariat du CNE, la publication des rapports et qui surtout s'assure de la conduite effective des évaluations. Le CNE manquait par exemple de visibilité, en particulier dans les médias. Et le contrôle de qualité semblait dispersé entre plusieurs instances.

Pourtant d'après S. Le Bouler, les membres des différentes instances d'évaluation, qui sont mobilisés pour 12 à 18 mois pour chaque évaluation et dont la composition est faite d'un tiers d'experts et d'universitaires, d'un tiers de représentants des administrations et d'un tiers de membres de la société civile, se déclaraient satisfaits du fonctionnement du dispositif. S. Le Bouler y voit notamment une confirmation des vertus du modèle participatif.

Au total, si on devait tirer les leçons de cette expérience pour mieux définir les missions d'une évaluation au niveau interministériel, quelles missions retiendrait-on ? S. Le Bouler en distingue trois :

un rôle d'impulsion pour amener de nouveaux acteurs dans le secteur de l'évaluation et s'assurer que certains thèmes « orphelins » soient bien pris en compte,

une mission transversale limitée compte tenu des précédents échecs des procédures interministérielles,

un rôle de support, insuffisamment développé dans les expériences précédentes et pourtant fréquent à l'étranger, c'est-à-dire de moins en moins procéder à la réalisation concrète des évaluations et plus accentuer l'action du CNE ou d'une instance équiva-

lente sur la formation, la définition méthodologique de protocoles d'évaluation, la mise en réseau des ressources etc.

Quels en sont les enjeux ? D'une part il est important de développer la compétence stratégique au niveau central en se dégageant d'une vision héroïque de l'évaluation. D'autre part, il est nécessaire de faire le lien entre ce dispositif d'évaluation et la LOLF², c'est d'autant plus nécessaire pour S. Le Bouler que très peu d'actions sont faites en la matière pour le moment. Enfin la décentralisation impose de mieux cordonner ces évaluations au niveau national.

En guise de conclusion, S. Le Bouler cite Pierre Moscovici en charge du dispositif interministériel d'évaluation en 1994 :

« Au-delà des enjeux bureaucratiques bien minces des querelles de boutique en l'occurrence aussi courtoises que sont petites les échoppes, je suis convaincu que la politique à l'orée du XXI^e siècle saura faire sa place à cette discipline essentielle qu'est l'évaluation des politiques publiques. »

DÉBAT AVEC L'ASSEMBLÉE

Quel est l'impact des évaluations menées par la CEPP ? N'assiste-t-on pas à une instrumentalisation par les hommes politiques des conclusions des différents rapports dans le but de légitimer leur action ?

J.-D. Delley, volontairement, ne répond pas à la question de l'impact d'un point de vue statistique car pour lui la statistique n'est pas révélatrice. Il préfère donner un sentiment plus personnel. Pour le milicien qu'il a été pendant 8 ans, milicien mal payé selon lui, la seule satisfaction de son métier fut de voir bouger les choses. Il a pu observer que les résultats de son action tombaient régulièrement, même si cela prenait parfois 3 ou 4 ans. Toutefois il reconnaît volontiers les limites des travaux de la CEPP qui sont celles de toute évaluation : l'homme politique est toujours libre ou non d'en faire usage.

² Loi organique relative aux lois de finances.

« *On ne fait pas une évaluation contre un homme politique et son administration, ce n'est pas possible* » déclare-t-il.

Il insiste en revanche sur le faible degré d'instrumentalisation des évaluateurs par les hommes politiques genevois, et ce pour deux raisons principales. La première est que la CEPP s'est auto saisie de 80 % de ses mandats et a été entièrement libre de la rédaction des rapports d'évaluation des 20 % restants. La seconde concerne la publication des résultats. La publicité des résultats est essentielle à une évaluation digne de ce nom et au fonctionnement démocratique d'une société. L'évaluation peut être critiquée, mais dans un débat public. La CEPP a donc entièrement pris en charge la publication de ses rapports pour ne pas la laisser aux hommes politiques. Contrairement aux universitaires qui ont tendance à croire que la simple publication suffit pour que leurs idées s'imposent, les hommes politiques tentent toujours de prendre en charge la publication et d'imposer leur propre communication sur un sujet. C'est pourquoi la CEPP a dès le début organisé des journées entières de communication avec la presse. Pour J.-D. Delley, l'évaluateur doit aller jusqu'au bout et prendre le temps de dire à la population ses conclusions. C'est comme cela qu'il peut s'émanciper du carcan établi en amont par le commanditaire. « En montrant le miroir, on montre un carcan fissuré », déclare-t-il.

Il a bien conscience toutefois que beaucoup de choses se jouent en amont. Certains hommes politiques essaient d'orienter l'auto saisine de la CEPP. Ou alors ils confient à cette dernière, via le Parlement, des évaluations aux enjeux éminemment politiques comme celles sur les travailleurs clandestins par exemple. Ils tentent aussi parfois d'externaliser le coût de production de certaines informations vers la CEPP. Il est en effet commode pour un homme politique de confier à la CEPP, qui a un budget propre, une évaluation qui produira une quantité d'informations qui lui seront utiles, plutôt que de la confier à son administration et d'en assumer le financement. Il existe donc bien un dialogue décisif en amont entre commanditaires éventuels, administrations et évaluateurs, mais J.-D. Delley persiste à n'y voir qu'un dialogue utile à la démocratie genevoise, même s'il concède qu'il s'agit là d'un particularisme. « Cela marche peut-être parce que c'est un petit milieu » conclut-il.

LES ÉVALUATEURS SONT-ILS MAL PAYÉS ?

J.-D. Delley constate que la part du budget général consacré à la rémunération des miliciens est insuffisante pour que la CEPP mène à bien ses activités. La CEPP a par exemple de plus en plus de mal à recruter du personnel venant des entreprises privées. Ce personnel, bien payé généralement dans son secteur, est de moins en moins incité à donner de son temps à l'État qui ne le rémunère pas à la hauteur de ce que lui coûte le changement temporaire de fonction. M. Le Bouler constate à ce sujet que beaucoup des membres des instances d'évaluation coordonnées par le CNE étaient des bénévoles, mais ils passaient des commandes à des évaluateurs privés qui, eux, étaient très bien payés, ce qui rend la pérennisation du système en place bien évidemment difficile...

Quels liens peut-on établir entre les évaluations et la LOLF ? S'est-on appuyé sur ce que nous avaient appris les évaluations pour mettre en oeuvre cette nouvelle loi ?

S. Le Bouler apporte plusieurs éléments de réponse. Le mot « évaluation » apparaît dans l'article 7 de la Loi organique sur les lois de finances du 1^{er} août 2001 :

« Un programme regroupe les crédits destinés à mettre en œuvre une action ou un ensemble cohérent d'actions relevant d'un même ministère et auquel sont associés des objectifs précis définis en fonction de finalités d'intérêt général ainsi que des résultats attendus et faisant l'objet d'une évaluation. »

Un lien est fait a priori entre la gestion par les performances et l'évaluation. Toutefois ce lien n'est pas forcément substantiel et la pratique depuis 2001 ne plaide pas pour une pleine intégration des évaluations dans la LOLF. Ceux qui ont mis en œuvre la LOLF ont surtout retenu la gestion des performances sur la base d'indicateurs, et non en fonction d'un volet évaluation. On n'a jamais envisagé, selon lui, qu'au moment de fabriquer des programmes et des indicateurs, on puisse utiliser les évaluations existantes pour questionner certains liens de causalité ou même les fameux « effets propres » souvent cités dans la

loi. On a ainsi évacué les acquis potentiels des sciences sociales dans ces travaux préliminaires, les enquêtes ne devant pas coûter trop cher et produire des résultats dans un temps court. Ainsi, non seulement on n'a pas utilisé l'évaluation pour dessiner la nouvelle maquette budgétaire mais en plus, on l'a réduite à une somme d'indicateurs statistiques.

LE RECOURS À DES ÉVALUATEURS EXTERNES EST-IL FORCÉMENT UN GAGE DE QUALITÉ ?

S. Le Bouler prend l'exemple du rôle des fonds structurels européens dans le développement d'évaluations externes, souvent purement quantitatives, effectuées par les administrations territoriales. Pour lui, ces évaluations présentent deux défauts. D'une part, même si l'on observe le regroupement de certains acteurs, ces évaluations sont souvent externalisées vers des opérateurs qui cumulent plusieurs évaluations et qui ne peuvent pas toujours investir dans leur relation avec les évalués. D'autre part, cette absence de véritable participation des évalués n'est pas compensée par la mise en réseaux des ressources, par la capitalisation des acquis par les évalués notamment. S. Le Bouler reste donc sceptique quant au gain en maturité du recours à des évaluations externes dans les régions.

DEUXIÈME PARTIE

LES DONNÉES, MÉTHODES
ET DÉMARCHES D'ÉVALUATION

La seconde partie du séminaire proposait d'aborder l'évaluation comme enjeu de méthode. Bien évidemment, cette dimension est en forte interaction avec la première, à savoir l'enjeu politique de l'évaluation, pour deux raisons principales. La première raison tient au fait que c'est principalement sur la base de la méthodologie retenue que se fondent la plupart des débats ultérieurs sur l'évaluation, que s'apprécie sa pertinence et que se décide ou non son éventuel usage. La méthode permet alors à l'évaluation d'être un moyen puissant de légitimation d'une action politique, celle-ci pouvant se parer de l'objectivité scientifique, de l'efficacité technique ou de la représentativité citoyenne. À l'inverse, elle peut être aussi le prétexte à un déni total des conclusions de l'enquête, parfois sous couvert du bon sens, souvent parce que la dite méthodologie minore ou néglige tel ou tel aspect qui avait en général été mis en avant par les acteurs désireux de participer à la construction du diagnostic.

La seconde raison renvoie à la méthode elle-même. Si la méthode peut être considérée comme « la matière vivante » de l'évaluation, elle n'en est pas pour autant totalement neutre sur le plan politique. Selon le degré d'ouverture et de dialogue entre les participants qu'elle permet, et selon le degré de (re)construction possible de l'objet de l'évaluation qu'elle autorise, elle n'intervient pas de la même manière et au même moment dans le cycle de production d'une politique publique (Lascoumes, Setbon 1996). Pur instrument de légitimation ex-ante d'une réforme dans un cas, elle peut permettre de redéfinir un problème public dans l'autre, ou simplement contribuer à encadrer le débat public¹.

Enfin, et peut-être surtout, le choix d'une méthode détermine en partie les conclusions auxquelles peuvent aboutir les évaluateurs, et donc leurs consé-

¹ C'est la position défendue par C. Thélot par exemple, ancien directeur du HCEE (Haut Conseil de l'évaluation de l'école) et de la DEP (Direction de l'évaluation et de la prospective) au ministère de l'éducation nationale. Pour lui l'évaluation a pour principale fonction de structurer le débat foisonnant sur l'école en France en produisant ce qu'il appelle un « effet miroir » auprès des acteurs. Voir notamment C. Thélot, « À quoi sert l'évaluation des politiques publiques ? », in M.-C. Kessler, P. Lascoumes et J.-C. Thoenig (dir.), *Évaluation des politiques publiques*, Paris, L'Harmattan, 1998.

quences politiques. Deux exemples rapides, pris dans le domaine des politiques éducatives, illustrent cette situation. Une récente étude de T. Piketty², sur l'effet de la réduction de la taille des classes sur la réussite scolaire, remet en cause bien des conclusions acquises sur le sujet alors qu'elles se fondent sur des données déjà connues et rassemblées par la DEP³. Une méthode statistique nouvelle, qui prend en compte les seuils à partir desquels on ouvre ou on ferme des classes dans différentes académies, lui permet en effet de conclure, à l'opposé des principales enquêtes sur le sujet, à un effet important de la réduction du nombre d'élèves par classe sur la réussite scolaire, et même à prédire que cet effet serait bien plus grand si on le ciblait sur certaines catégories d'élèves. On voit bien ici comment une innovation statistique, donc méthodologique, peut relancer la question éminemment politique des « moyens » en éducation. Le second exemple concerne l'évaluation des politiques de lutte contre les inégalités à l'école. Un récent article de J.-C. Combessie au titre évocateur⁴ montre en effet que dans une situation particulière, mais très fréquente dans le système éducatif⁵, il existe autant d'applications statistiques qui permettent de conclure à une réduction des inégalités qu'à leur maintien, voire leur accroissement. D'après lui, il n'existe aucune raison mathématique de préférer le recours à une application plutôt qu'à une autre. Pourtant une rapide étude des principaux ouvrages sur l'état des inégalités à l'école révèle

² Étude disponible en ligne à l'adresse suivante :

<http://www.jourdan.ens.fr/piketty/fichiers/public/Piketty2004b.pdf>
<http://www.jourdan.ens.fr/piketty/fichiers/public/Piketty2004b.pdf>

Voir également l'article publié dans *Le Monde* du 6 septembre 2004.

³ Direction de l'évaluation et de la prospective.

⁴ J.-C. Combessie, « Trente ans de comparaison des inégalités des chances. Quand la méthode retenue conditionne la conclusion », *Courrier des statistiques*, n° 112, 2004.

⁵ Cette situation se caractérise par le fait que les termes d'une inégalité varient dans le même sens ($A_1 < A_2$ et $B_1 < B_2$) et que le sens de l'inégalité initiale est conservée ($A_1 > B_1$ et $A_2 > B_2$). Pour prendre un exemple, A_i pourrait désigner le taux d'admission en terminale S des enfants de familles privilégiées à la période i et B_i celui des enfants de familles défavorisées. Dans les deux cas le taux d'admission augmente entre les deux périodes, mais celui des enfants de familles favorisées reste supérieur pour les deux périodes considérées.

que les auteurs optent dans la grande majorité des cas pour des applications statistiques permettant de conclure à une baisse des inégalités. Il s'inquiète alors du développement d'une relative orthodoxie en statistique qui voudrait que certaines applications (odds ratios et taux logistique notamment) soient préférables aux autres alors que toutes sont également fondées sur le plan mathématique. Cet article montre que le choix de la méthode peut aboutir (ou non) à « *une représentation enchantée de l'évolution des inégalités* » (p.1) pouvant justifier (ou non) l'existence même d'une politique de lutte contre les inégalités à l'école.

Tout analyste sérieux de l'évaluation ne peut donc se dispenser d'une plongée dans la « cuisine méthodologique » des évaluateurs. Mais quels sont les ingrédients de ce processus et quels en sont les enjeux ? Dans un premier temps, S. Le Minez et M. Lelièvre font part de leur expérience professionnelle de statisticiennes au sein du ministère des affaires sociales, et en particulier de la DREES⁶ et posent la question épineuse de la construction d'indicateurs solides, au niveau européen notamment, dans des domaines aussi complexes et pluridimensionnels que la pauvreté ou l'emploi. J.-L. Richard quant à lui, aborde les problèmes tant politiques que scientifiques, voire éthiques, de la construction et de l'utilisation en France de statistiques publiques de la nationalité pour lutter contre les discriminations. B. Rihoux enfin évoque les possibilités offertes à l'évaluation par le recours à une méthode particulière : l'analyse quali-quantitative comparée (AQQC).

En guise de synthèse, G. Hollard et A. Vion fournissent une grille de lecture générale des différentes méthodologies utilisées. Sur la base de deux variables, le degré de prise en compte de l'information par les évaluateurs et le degré de hiérarchisation de cette même information, les auteurs distinguent quatre types d'évaluation (utilitariste, instrumentale, praxéologique et procédurale) qu'ils déclinent selon plusieurs catégories : le principe général, le style de démarche, la nature des expériences que l'évaluation génère, le type de modèles et d'enquêtes construits et les enjeux des débats qu'elle soulève.

⁶ Direction de la Recherche des études et de l'évaluation des statistiques.

I

CONSTRUCTION ET USAGES DES BASES DE DONNÉES ET DES INDICATEURS STATISTIQUES

ÉLÉMENTS DE RÉFLEXION SUR LA CONSTRUCTION, LES USAGES ET LA PORTÉE DES INDICATEURS STATISTIQUES : UNE EXPÉRIENCE PROFESSIONNELLE DANS UNE ADMINISTRATION SOCIALE COMME LA DREES

Sylvie LE MINEZ et Michèle LELIEVRE
ministère des Affaires sociales, DREES

Les auteures présentent leur expérience professionnelle au sein de la Direction de la recherche des études et de l'évaluation des statistiques du ministère des Affaires sociales en France. Cette expérience est différente pour chacune d'elle. S. Le Minez s'occupe du pan évaluation et est chargée en particulier de travailler sur des micro-simulations, c'est-à-dire des modèles qui permettent de tester l'effet des transferts sociaux sur les revenus des ménages et donc de simuler différentes réformes. Le travail de M. Lelièvre est quant à lui plus

centré sur la construction et l'usage d'indicateurs statistiques dans des domaines variés comme les retraites, l'exclusion sociale ou l'Union européenne.

S. Le Minez, après avoir cité un certain nombre de sources sur lesquelles repose son intervention¹, rappelle quelques éléments de contexte sur l'apparition des indicateurs sociaux en France et dans d'autres pays. Dès les années 1970 apparaissent les premières réflexions autour de ces indicateurs. L'optique est alors de produire d'autres statistiques pour compléter ou contrebalancer la vision du développement économique donnée par certains indicateurs économiques dominants comme le PIB. Ces réflexions ont débouché sur la mise en place de comptes satellites comme les comptes de la protection sociale recueillis chaque année par la DREES et qui deviennent de véritables outils de pilotage. Aujourd'hui un nouvel élan est donné à la construction de ce type d'indicateurs dans un contexte de construction européenne, de mise en place de politiques d'évaluation et également de la LOLF.

Qu'est-ce alors qu'un indicateur social ? C'est avant tout une statistique choisie pour sa capacité à agréger de façon optimale les informations relevant de phénomènes sociaux complexes. Les objectifs assignés à ces indicateurs peuvent être multiples : rassembler l'information, permettre l'évaluation d'une politique ou la comparaison de situations nationales par exemple. Quelles sont les principales qualités d'un indicateur ?

- Il doit être univoque pour éviter au maximum les ambiguïtés et les mauvaises interprétations à son sujet,
- il doit être le plus représentatif possible,

¹ Comme par exemple le rapport de B. Perret, *Indicateurs sociaux, état des lieux et perspective*, établi pour le Conseil de l'emploi, des revenus et de la cohésion en janvier 2002 ; les travaux de l'INSEE sur la pauvreté monétaire relative et les autres formes de pauvreté (notamment les travaux de S. Lollivier, et de D. Verger) ; la charte des indicateurs de T. Atkinson ; les travaux de la mission interministérielle de la recherche sur la mesure des inégalités (l'article de C. Chambaz notamment) ; ou encore certaines publications de la DREES comme celles de M. Lelièvre sur les indicateurs européens.

- il doit témoigner d'une certaine « clarté normative », c'est-à-dire qu'on doit être capable de lui assigner des valeurs (positives ou négatives) claires pour procéder à des agrégations de données,
- il doit être robuste, juste, transparent, et accessible.

Faut-il avoir recours à un indicateur unique ou à plusieurs indicateurs ? S. Le Minez nous met en garde contre l'usage de multiples indicateurs. Certes cet usage permet de multiplier les approches d'un même phénomène complexe. Mais se posent alors d'autres difficultés. Comment pondérer ensuite les différents indicateurs ? Qui doit procéder à cette pondération ? La plupart du temps ce sont les experts qui la réalisent, mais ce choix n'est-il pas éminemment politique ? Enfin l'interprétation de l'évolution de cet indicateur synthétique final est souvent ardue, alors qu'il a par ailleurs un pouvoir médiatique et symbolique indéniable. C'est pourquoi il faut veiller à ce que les indicateurs ne soient pas trop nombreux, qu'ils se complètent ou qu'ils s'équilibrent et qu'ils soient hiérarchisés, sinon l'information est trop importante et l'interprétation impossible. C'est tout l'enjeu de la définition d'un taux de pauvreté monétaire relative au niveau européen.

L'évaluation à la DREES ne se limite pas à l'exploitation d'indicateurs sociaux mais se fonde également sur d'autres outils : les comptes de la protection sociale déjà évoqués (part des dépenses en protection sociale dans le PIB etc.), les comptes administratifs (donnant par exemple le nombre de bénéficiaires de telle mesure), les enquêtes statistiques (panels, cohortes...) et des modèles d'analyse conjoncturelle (batteries d'indicateurs, modèles macroéconomiques, modèles de simulation etc.). Ainsi l'évaluation du dispositif de la prime pour l'emploi a nécessité la mobilisation de toute une palette d'outils pour pouvoir en cerner tous les aspects.

La DREES a, entre autres, des missions d'évaluation particulières au niveau européen. Elle pilote par exemple la délégation française du ISG-SPC, (sous-groupe indicateurs du Comité de la protection sociale) à Bruxelles, rattachée au comité de protection social dépendant lui-même du Conseil européen (depuis le Conseil européen de Lisbonne de mars 2000). Ce comité a reçu pour mandat de mettre en œuvre la méthode ouverte de coordination dans le

secteur social et plus récemment dans le champ de la santé. Ce sous-groupe technique a été mandaté par le Conseil européen pour choisir une batterie d'indicateurs représentatifs des objectifs que se sont fixés en commun les différents États membres.

Deux contraintes importantes ont précédé le choix de ce benchmark : d'une part les acteurs de la DREES ont dû trouver un compromis dans un temps court dans le champ retenu ; d'autre part, le choix des indicateurs a été également limité, notamment en ce qui concerne les indicateurs communautaires (comme le panel des ménages européens géré par différents instituts statistiques nationaux ou comme le système statistique européen harmonisé des comptes de la protection sociale). Quels sont les principes fondateurs ayant présidé à ce choix ? Quel était l'état du consensus sur ce sujet ?

L'indicateur doit répondre à la Charte des indicateurs statistiques établie par un comité d'experts présidé par T. Atkinson. Il doit être accessible, attester d'une certaine robustesse statistique et ne pas être sujet à manipulation. Pour des raisons politiques, les États membres se sont accordés pour ne retenir que des indicateurs de résultat, qui sont essentiellement descriptifs, qui ne décrivent que l'impact final après intervention publique et qui ne permettent pas de rendre compte de la manière dont est solutionné le problème social en question. L'indicateur doit rarement être utilisé seul. La pratique, et non les textes, montre qu'il doit être accompagné de variables contextuelles (souvent démographiques) dans le but de saisir l'aspect multidimensionnel de la pauvreté. Cette dernière fut saisie de trois manières : dans ses aspects multidimensionnels par une batterie d'indicateurs, mais aussi à travers une définition de la pauvreté absolue et des indicateurs de condition de vie. Enfin l'usage d'indicateurs composites a été écarté pour limiter les dimensions d'un phénomène déjà complexe.

La majorité des indicateurs sont définis de façon communautaire. Il y a comme toujours des exceptions comme la définition d'un logement décent par exemple, qui dépend du niveau de vie du pays considéré. Une vingtaine d'indicateurs sont retenus (les indicateurs de Laeken) selon une architecture à trois niveaux : le premier niveau regroupe des indicateurs reconnus et jugés

primordiaux par un grand nombre d'experts, le deuxième niveau des indicateurs de complément du premier niveau et le troisième des indicateurs qui ne font pas consensus.

Ces indicateurs font l'objet de plusieurs critiques. Ils sont très globaux et il est difficile de les relier à un volet particulier de la politique sociale, bien qu'ils permettent tout de même d'évoquer des problèmes de redistribution au niveau global. Le traitement longitudinal de ces indicateurs est encore perfectible. Ils ne sont pas immédiatement disponibles (les dernières données disponibles datent par exemple de 2000). Enfin ils ne sont pas toujours harmonisés d'un pays à l'autre. Par exemple on ne prend pas en compte en général dans le calcul du pouvoir d'achat le pouvoir d'achat en retour que permet la fourniture de biens publics. Or les pays scandinaves intègrent cette dimension dans leurs calculs.

Le sous-groupe « indicateurs » est pour le moment confronté à trois types de difficultés : comment déterminer un indicateur européen décrivant la situation d'un travailleur pauvre ? Comment se positionner face à la méthodologie utilisée par l'OCDE et le comité de politique économique pour élaborer des indicateurs du degré plus ou moins rémunérateur du travail ? Comment construire des indicateurs de l'état de la redistribution et notamment de l'adéquation des pensions versées ? La France par exemple refuse les indicateurs de niveau de vie relatif quand il s'agit d'évaluer le niveau de vie des personnes âgées permis par leurs pensions. Elle ne les juge pas révélateurs des États Providence bismarckiens, alors que le Royaume-Uni milite pour l'utilisation de ces indicateurs. Ces querelles montrent, les intervenantes insistent sur ce point, que les indicateurs ne sont jamais neutres, qu'ils renvoient toujours à une théorie économique sous-jacente ainsi qu'à une théorie particulière de la justice sociale.

LA STATISTIQUE PUBLIQUE
DE LA NATIONALITÉ
ET DES ORIGINES CULTURELLES
ET GÉOGRAPHIQUES
À L'HEURE DES POLITIQUES
PUBLIQUES DE LUTTE
CONTRE LES DISCRIMINATIONS

Jean-Luc RICHARD
CRAPE UMR CNRS

J.-L. Richard rappelle l'actualité de la question. En 1997 est formé le Groupe d'étude et de lutte contre les discriminations (GELD) par M. Aubry avec la création notamment d'un numéro vert, le 114, que tout individu peut composer s'il s'estime victime de discrimination. L'auteur voit en cette date le commencement d'une politique explicite de lutte contre les discriminations, les politiques précédentes traitant plutôt de la baisse des inégalités¹ ou de l'intégration comme en témoigne la création à la fin des années 1980 du Haut Conseil à l'intégration, avec son groupe chargé de la production de statistiques. L'article 13 du Traité d'Amsterdam donne lieu en juin 2000 à une directive sur la lutte contre la discrimination, adoptée tardivement par la France notamment parce que les termes de « race » et d'« ethnie » y apparaissaient. Sur cette base fut votée en France en novembre 2001 une loi transcrivant la

1 Même si cette lutte contre les inégalités se voulaient parfois « active » comme dans les cas de la politique des ZEP (Zone d'éducation prioritaire) ou de celle des ZUS (Zones urbaines sensibles).

directive et modifiant la loi de 1972. Elle inverse la charge de la preuve : c'est à celui qui est accusé de discrimination de prouver qu'il ne l'a pas pratiquée. D'où l'évolution du répertoire d'action de certains mouvements sociaux comme l'illustre la pratique du « testing » développée par SOS Racisme. D'où également l'évolution des recherches en sciences sociales sur le sujet, recherches qui incorporaient plus ou moins des recherches quantitatives.

Contrairement à certaines idées reçues sur le cas français², depuis 1862, on distingue en France parmi l'ensemble des Français ceux qui sont nés de parents étrangers, et on dispose même de statistiques sur la nationalité antérieure de ces Français « par acquisition ». On dispose également de statistiques sur les enfants de ces personnes, donc sur les enfants d'immigrés (immigrés définis par le HCI et la communauté des démographes comme des personnes nées étrangères hors de France³). Ces statistiques permettent notamment d'étudier le chômage ou la réussite scolaire des personnes étrangères, des immigrés et de leurs enfants (les personnes dites « issues » de l'immigration).

Ces statistiques proviennent de sources diverses : les enquêtes FQP (Formation et qualification professionnelle), les enquêtes du CEREQ sur l'entrée dans la vie active, le recensement, l'échantillon démographique permanent issu du recensement qui permet des analyses longitudinales et l'enquête « famille » de l'INSEE, appelée maintenant « histoire de vie ». Ces statistiques permettent de montrer par exemple que les jeunes d'origine algérienne sont plus souvent au chômage que ceux d'origine marocaine ou tunisienne, ou même ceux venant d'Afrique subsaharienne, eux-mêmes bien plus au chômage que le reste des jeunes français. Une enquête reste cependant à faire sur le rôle discriminant de l'école, rôle discriminant qui ne s'observe pas a priori en termes de niveau d'étude atteint mais en termes d'orientation et de choix de filières.

2 Dont on dit rapidement qu'il ne produit aucune statistique sur les immigrés et leurs enfants sous prétexte qu'ils sont soit « français » et intégrés au modèle républicain soit étrangers.

3 L'ONU parle simplement de « Foreign Born », c'est-à-dire de personnes nées en dehors du pays. Avec cette définition, la France par exemple n'aurait pas 4 mais 6,5 millions d'immigrés.

Le débat qui secoue la France sur la légitimité de statistiques sur l'origine culturelle ne lui est pas particulier. À l'exception de la Grande-Bretagne où depuis 1991 il existe des statistiques dites « ethniques », aucun pays de l'Union européenne ne dispose de telles statistiques et tous ont un débat sur l'opportunité ou non d'en créer⁴.

Tout l'enjeu est alors de savoir si ces données sont suffisantes pour produire des connaissances fiables sur le sujet, dont les citoyens pourraient s'emparer lors de leurs actions en justice. Pour J.-L. Richard en effet, la politique de lutte contre les discriminations se résume pour le moment essentiellement au recours au droit que permettent les lois de 1972 et de 2001.

C'est pourquoi certains chercheurs en appellent à la création d'une véritable statistique ethnique sur le sujet afin de pouvoir apporter des réponses claires. Cette statistique aurait le mérite de se fonder sur les identités déclarées des individus, et de ne pas leur imposer une ou plusieurs identités particulières. Ces chercheurs soulignent les avantages actuels des données rassemblées par les États-Unis sur le sujet. L'expérience a montré en outre que la récolte de ces données par l'Union européenne était facile et relativement fiable, 99 % des individus sondés acceptant de donner leur nationalité de naissance et celle de leurs parents. Enfin on peut considérer qu'il fait partie des prérogatives d'un État de connaître les populations qu'il intègre.

Néanmoins le recueil de ces données pose plusieurs problèmes. Le premier se résume à une question simple : comment dire d'un enfant qu'il est fils d'immigré(s) ou même petit-fils d'immigré(s) ? Est-ce qu'on est fils d'immigré avec trois grands-parents sur quatre qui sont immigrés par exemple ? Le deuxième problème est que ces données risquent de réifier certaines identités et d'alimenter une forme de communautarisme. Pour reprendre les mots de J.-L. Richard, « On va mettre dans une case ceux qui se revendiquent comme tels et peut-être passer à côté de ceux qui ne revendiquent pas pour plusieurs

⁴ Ce qui n'empêche pas la reconstruction a posteriori de ces catégories ethniques, parfois dans des travaux de sciences sociales comme a pu le remarquer J.-L. Richard dans le cas des Pays-Bas, pays qui utilise au départ la même nomenclature que la France (individus étrangers, immigrés, fils d'immigrés...)

raisons mais qui eux aussi souffrent de discrimination. » On voit d'ailleurs que ces catégories sont de plus en plus remises en cause aux États-Unis et que par exemple le nombre d'Hispaniques varie fortement selon que la catégorie est placée en première position ou non. Ces statistiques peuvent également être facilement instrumentalisées d'un point de vue politique.

Enfin quand bien même ces données seraient-elles produites, il n'est pas certain qu'elles permettent de montrer qu'il y a ou non discrimination. Les chercheurs en sciences sociales sont d'ailleurs divisés sur le sujet. Pour certains il faut démontrer l'intentionnalité du discriminant pour démontrer l'existence d'une discrimination. Pour d'autres, une démonstration indirecte est possible et peut suffire. C'est dans ce dernier courant que s'inscrit pour l'instant la majorité des chercheurs en économie du travail et en sociologie pour J.-L. Richard, à l'exception par exemple des travaux actuels de P. Amadiou (Paris I).

DÉBAT AVEC L'ASSEMBLÉE

Intervention d'Alain Chenu (OSC-CNRS), discutant

A. Chenu introduit les débats en soulignant deux points communs aux interventions de S. Le Minez, de M. Lelièvre et de J.-L. Richard. D'une part il remarque que ce qui fait un bon indicateur, c'est toujours le cumul de la qualité méthodologique et de la pertinence du sens. Toutefois cette pertinence varie selon les contextes, ce qui explique que le degré de consensus sur la classification des indicateurs varie lui aussi. Les débats autour du seuil de pauvreté en sont un bon exemple. Le calcule-t-on de façon uniforme au niveau européen ou garde-t-on les seuils nationaux ? A. Chenu prend également l'exemple de la catégorie « juif » dans l'Allemagne de l'entre-deux guerres. Cette catégorie demeure une catégorie religieuse dans les années 1920 alors qu'elle devient dans les années 1930 une sous-catégorie de la catégorie « race ». Il en conclut qu'il existe une corrélation entre le degré de consensus idéologique de la communauté scientifique et la définition de ce qu'est une construction méthodologique solide des indicateurs pour cette même communauté. Les scien-

tifiques tombent d'autant plus d'accord sur la manière dont l'indicateur est « encapsulé », pour reprendre l'expression de A. Desrosières, qu'il existe parmi eux un consensus politique ou idéologique. Il constate enfin qu'on retrouve toujours des questions sensibles autour de la construction des indicateurs et que selon l'ampleur du détour que doit faire le statisticien pour produire sa statistique, il perd plus ou moins en pertinence. L'ampleur de ce détour dépend pour A. Chenu du degré d'implication du citoyen dans le problème concerné.

D'autre part, A. Chenu pose la question de la méta-évaluation, c'est-à-dire de l'appréciation des indicateurs supposés aider à l'évaluation. Il constate que dans les deux domaines évoqués plus haut, cette appréciation est laissée aux seuls spécialistes de la statistique et qu'il se pose également souvent un problème d'accès aux sources. La plupart des travaux sur le recensement, comme l'échantillon permanent tiré du recensement, réalisés par l'INSEE par exemple ne sont en général disponibles que depuis les bureaux de l'INSEE.

Qui décide de la pertinence des indicateurs statistiques retenus ? N'est-ce pas une décision éminemment politique ?

M. Lelièvre précise le partage « naturel » des tâches entre expertise et politique tel qu'il est actuellement pratiqué : les objectifs sont fixés par les hommes politiques, les benchmarks pour atteindre ces objectifs sont formulés par les experts, qui sont en grande partie des « policy unit » croisant les deux compétences (politique et technique). Elle fait remarquer que ce sont souvent les statisticiens eux-mêmes qui mettent en garde les hauts-fonctionnaires contre un usage trop rapide, global et unique de leurs indicateurs. La qualité du dialogue dépend ensuite de la composition des directions centrales des ministères. En Italie par exemple, les membres des directions centrales sont en majorité des économistes.

Le rôle joué par la DREES dans la définition des indicateurs retenus par la LOLF dans le domaine de la santé est un autre bon exemple. La DREES a bien été associée à leur définition. Mais elle a dû dialoguer avec des acteurs porteurs d'autres cultures administratives. Que faire par exemple lorsque l'on

constate une diminution du montant d'une allocation différentielle comme l'allocation « parent isolé » par exemple ? Faut-il conclure rapidement comme certains le font à une amélioration de la situation globale justifiant un moindre recours à ce type d'allocation ? Ou faut-il, comme M. Lelièvre le souhaiterait, vérifier si on n'a pas coupé dans les budgets de cette allocation l'année précédente ?

Concernant les indicateurs de résultat retenus par l'Union européenne, la décision fut prise par une majorité de délégations nationales, le rôle de la Commission étant inégal selon les comités. M. Lelièvre note par exemple qu'elle fut en retrait de son comité et plus globalement de ceux qui étaient regroupés sous le pôle emploi.

C. Reuters (attachée scientifique à l'Institut wallon de recherche pour l'évaluation et la prospective de statistique) fait part d'une expérience menée en Belgique, où un groupe de travail fut institué par un représentant belge au niveau européen pour discuter de la pertinence des indicateurs retenus pour mesurer la pauvreté. Ce groupe était composé d'évaluateurs et de personnes en situation de pauvreté et fut traversé de nombreux débats. Il a produit un rapport dont le but était d'aider à penser la pauvreté au niveau de la Belgique et qui a eu pour effet d'élargir la batterie d'indicateurs.

L'Union européenne ferait-elle le choix de ne pas évaluer ses politiques en ne retenant pas d'indicateurs de processus mais que des indicateurs de résultat ?

M. Lelièvre reconnaît volontiers qu'il existe des contradictions dans le fonctionnement de l'Union européenne. Au niveau strictement communautaire, les délégations s'accordent à peine sur des indicateurs communs permettant une simple description des phénomènes. Mais elles ne sont pas toutes d'accord pour procéder à des benchmarks et identifier les meilleures méthodes. Ce qui peut apparaître contradictoire avec la méthode ouverte de coordination, qui prévoit au final un « peer review » dont le but explicite est d'identifier et de diffuser les meilleures politiques publiques.

Quelles sont la fiabilité et la pertinence des micro simulations pour aider à la formulation d'une politique publique ? Celle-ci doit-elle forcément agir sur les comportements ?

S. Le Minez concède que de plus en plus de politiques économiques ou sociales visent à changer les comportements. La prime pour l'emploi, la réforme de l'assurance maladie ou de l'assurance chômage sont pour elle de bons exemples. Elle remarque toutefois que la France a beaucoup moins recours à ce type d'outil que la Grande-Bretagne par exemple. Cette technique des micro simulations consiste à partir d'un échantillon de ménages dont on connaît les revenus primaires et la composition et à le perturber par l'introduction de différentes réformes des revenus de transfert pour prévoir l'évolution de leur niveau de vie. Le recours à ces modèles permet donc d'évaluer ex-ante une politique sur la base d'hypothèses a priori. Mais ces simulations restent statiques et il manque encore des informations sur l'évolution du comportement des agents dont les préférences sont stabilisées. Changent-ils de comportement pendant la réforme des revenus de transfert ? Par exemple il est possible de modéliser l'offre de travail mais qu'en est-il de la demande de travail ? Il existe toutefois une autre solution qui consiste à observer l'évolution de certaines variables après la mise en œuvre d'une politique (taux d'emploi, pouvoir d'achat...) à l'aide de modèles macroéconomiques complexes pour l'évaluer a posteriori.

Pourquoi dévie-t-on souvent vers des considérations purement statistiques quand on parle d'évaluation ?

P. Duran (GAPP-ENS Cachan) propose des éléments de réponse. Pour lui le rapport Deleau⁵ est une merveille de naïveté épistémologique puisqu'il suppose 1/ que l'évaluation mesure des effets et 2/ que la vérité est unique⁶. Dès lors il n'est pas surprenant qu'on passe rapidement d'une intuition ou d'une

5 M. Deleau, *Évaluer les politiques publiques : méthode, déontologie, organisation*, Paris, La Documentation française, 1986.

6 P. Duran reconnaît que cette naïveté n'est pas explicite dans le rapport même si plusieurs passages font preuve de scientisme. Elle est plutôt le fait d'acteurs qui se sont ensuite saisis du rapport.

volonté politiques à une activité statistique car on y voit là une méthode d'action efficace. Mais cette méthode repose sur une épistémologie naïve car le propre des sciences sociales est de problématiser. Il y a donc forcément pluralisme méthodologique, la méthode étant fortement liée à la façon dont on interroge les phénomènes sociaux. Cette naïveté explique que certains soient toujours surpris, voire scandalisés, quand ils obtiennent des résultats contradictoires. Les indicateurs drainent toujours des arrière-plans. « Derrière les faits il y a toujours un peu de théorie » disait Goethe. Or les personnels administratifs ont tendance à prendre les données pour des faits. De même il existe toujours un effort de standardisation des évaluations, mais leur fiabilité méthodologique est souvent suspecte.

II

LES APPROCHES PLURIDISCIPLINAIRES DE L'ÉVALUATION

L'ANALYSE QUALI-QUANTITATIVE COMPARÉE (AQQC-QCA) : PROMESSES ET PLUS-VALUES POUR L'ÉVALUATION

Benoît RIHOUX
UCL

B. Rihoux présente une méthode qualitative de traitement d'un certain nombre de variables, l'AQQC¹, et s'interroge sur son éventuel apport à l'évaluation². Il commence par rappeler les quatre grands défis de l'évaluation,

¹ Pour plus de précision sur cette méthode, le lecteur peut de reporter aux ouvrages de G. De Meur et B. Rihoux, *L'analyse quali-quantitative comparée (AQQC-QCA : approche, techniques et applications en sciences humaines*, collab. S Yamasaki,, Louvain-la-Neuve, Academia-Bruylant, 2002 ; et de C. Ragin, *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies*, Berkeley, Los Angeles and London, University of California Press, 1987 ainsi qu'au numéro 1 (11) de 2004 de la *Revue internationale de politique comparée* consacré à « L'analyse comparative configurationnelle » et coordonné par l'auteur.

² Sur ce lien en particulier, l'auteur cite également B. Befani, « La Qualitative Comparative Analysis (QCA) e la valutazione basata sulla teoria : un connubio possibile », *Rassegna Italiana Di Valutazione*,

en se limitant volontairement à l'évaluation comprise comme l'évaluation des « outcomes », la mesure des effets concrets d'une politique :

1. établir la « outcome line », c'est-à-dire l'ensemble des effets attendus d'une politique sur les acteurs et leurs croyances, elle-même déterminée par l'identification de différents mécanismes de causalité ou par la formulation d'une théorie particulière du changement,
2. isoler les « effets nets », éliminer les « confounding factors »,
3. estimer le « policydeadweight », c'est-à-dire estimer dans une analyse contrefactuelle ce qu'il se serait produit si la politique n'avait pas été mise en œuvre,
4. trianguler les comparaisons (avant/après, avec/sans, selon les ères géographiques ou institutionnelles).

B. Rihoux défend avec C. Ragin³ l'idée qu'il est possible de trouver une méthodologie d'enquête fondée à la fois sur une analyse qualitative de cas et sur une analyse quantitative de variables. Il se fonde pour cela sur une démarche synthétique reposant sur l'algèbre booléenne : l'AQQC. Cette méthode est disponible dans plusieurs logiciels (fs/QCA; TOSMANA) dont certains sont gratuits.

Pour commencer, il faut dresser un tableau brut d'une vingtaine de variables significatives des cas retenus. Puis il faut dichotomiser⁴ les variables en un nombre limité de conditions (0 ou 1, bas et haut etc.), ce qui permet d'établir un tableau des configurations. On regroupe alors les situations très proches ou identiques pour qu'elles ne forment qu'une seule et même configuration. On

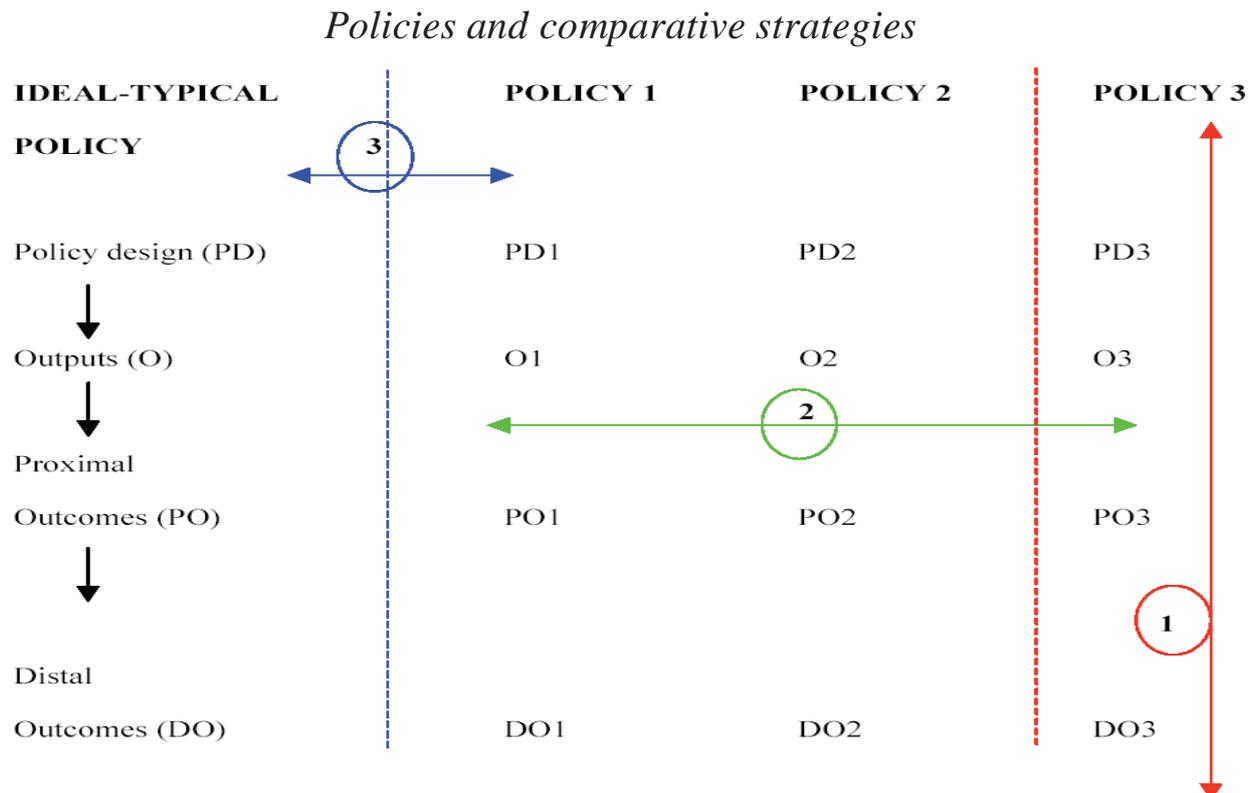
n° 2, 2004.

3 *Op. cit.*

4 Il existe des méthodes plus complexes dites de « Multivalue QCA » mais pour les besoins de la présentation, B. Rihoux n'entre pas volontairement dans ces détails.

minimise ensuite les données conformément à la procédure de minimisation booléenne pour obtenir une synthèse des données et des régularités. C'est l'opération centrale de l'analyse. On obtient alors la « formule minimale » que le chercheur doit ensuite interpréter.

Figure 1 : Exemple de formule minimale



Source : B. Rihoux, A. Marx, F. Varone (2005)⁵

Cette formule identifie différents chemins dans notre tableau général, différentes situations possibles qu'il convient d'interpréter. Trois utilisations de cette formule sont possibles :

⁵ B. Rihoux, A. Marx, F. Varone, « A new method for policy evaluation ? Longstanding Challenges and the Possibilities of Qualitative Comparative Analysis (QCA) », in B. Rihoux, H. Grimm (eds), *Innovative comparative methods for policy analysis and evaluation* (titre provisoire), New York Springer/Kluwer, 2005, à paraître.

une analyse verticale des occurrences qui correspond à une analyse riche au cas par cas (« within case analysis »),

une analyse horizontale des occurrences permettant une analyse systématique entre les cas (« systematic cross-cases analysis⁶ »),

une analyse en diagonale qui permet de comparer des séries de politiques par rapport à une politique publique idéal-typique, à condition que la théorie soit bien consolidée dans le secteur en question.

Qu'apporte cette méthode dite de l'AQQC à l'évaluation de politiques publiques ? Reprenons dans l'ordre les quatre défis évoqués précédemment :

1. l'AQQC n'intègre pas en tant que tels les processus même si elle repose sur une analyse fine des cas étudiés (« case based knowledge »). Mais il est possible d'interpréter la formule minimale en termes de processus et d'identifier les différents chemins menant au même résultat, ce qui est une véritable plus-value par rapport à d'autres méthodes statistiques,
2. elle permet la mise en œuvre de la méthode MSDO (« Most Similar/Different Outcome ») et d'évacuer certains cas. Elle permet aussi de regrouper les cas par grappes significatives et voir si telle variable macro ou tel élément de contexte joue ou ne joue pas⁷,
3. on peut sélectionner les cas non observés (les « remainers ») et les intégrer dans l'analyse,
4. l'algèbre booléenne, par définition, permet facilement de trianguler (« 011 », « 101 », « 011 »). Cette procédure de triangulation fut déjà appliquée par certains sociologues et politologues, initiateurs de la méthode, qui ont pu comparer les effets de l'État Providence aux États-Unis par exemple.

6 À ce sujet B. Rihoux conseille au lecteur de parcourir le site suivant : <http://www.compass.org>

7 Remarque : d'après B. Rihoux, l'AQQC ne permet pas de quantifier et ne permet de parler qu'en termes probabilistes.

B. Rihoux conclut en précisant que l'utilisation de cette méthode suppose à chaque étape une certaine transparence des choix (qui sont parfois de véritables dilemmes) et qu'elle permet de synthétiser les nombreuses études qualitatives existantes dans un domaine particulier⁸.

DISCUSSION

Interventions de P. Le Galès, C. Musselin et A. Van Zanten, discutants

A. Van Zanten (OSC-CNRS) reconnaît que cette démarche a le mérite de repousser plus loin les limites de la généralisation des approches qualitatives et d'être délibérément comparative, ce qui est d'autant plus intéressant dans un contexte scientifique où l'on manque d'études qualitatives et comparatives. Toutefois, elle émet trois réserves principales quant à l'utilisation d'une telle méthode.

La première est que, dans toute recherche, le choix des cas à étudier est le résultat d'une première comparaison, alors que l'AQQC donne l'impression que la comparaison vient après les études de cas, que celle-ci devient un but et non une stratégie de recherche. Or le sociologue qualitatif a besoin de revenir sur son terrain après avoir réalisé un premier temps de comparaison.

La deuxième est que l'application de cette méthode à l'analyse des politiques publiques, et encore plus à leur évaluation, n'est pas neutre sur le plan scientifique et repose sur un postulat implicite selon lequel la démarche de l'homme politique et de l'expert sont finalement proches, voire identiques et qu'on peut les retrouver par la méthode, l'AQQC. Or c'est trop minorer la dimension plus politique des politiques publiques. Comment réintégrer ensuite dans l'AQQC cette dimension fortement politique du contrôle et de la légitimation de l'action publique ?

⁸ Pour plus de précision le lecteur pourra se reporter au vol. 11, n° 1 de la *Revue internationale de politique comparée* de l'année 2004 qui présente 6 études de cas réalisées à partir de la méthode AQQC et fut coordonné par l'auteur.

La troisième concerne le choix des cas dont dépend grandement la validité de la méthode. Quand le chercheur est dans la recherche « pure », il construit une comparabilité. Quand on obéit à une commande politique, la comparabilité est donnée. On doit étudier les pays qui ont officiellement mis en œuvre telle politique, ce qui peut biaiser toute la méthodologie.

C. Musselin (CSO-CNRS) considère que dans cette analyse on ne pose jamais la question de la définition même d'un cas : qu'est-ce qu'un cas ? Qu'est-ce qu'une étude de cas ? De plus, tous les exemples pris sont finalement des situations dont les résultats sont quantifiables (nombre d'accidents de travail, taux d'accès à la propriété...). Comment alors procéder dans des cas essentiellement qualitatifs. Il est facile par exemple de déterminer combien d'universités sont passées au LMD, mais bien plus difficile de savoir si le LMD français est comparable au LMD belge (baccalauréat/master/doctorat).

Elle considère enfin que B. Rihoux développe une vision démesurément « balistique » des politiques publiques selon laquelle toutes les étapes d'une politique s'enchaînent clairement et fidèlement, alors que les problèmes ne sont pas figés au moment de lancer une politique et qu'ils évoluent pendant sa mise en œuvre. Comment intégrer cela dans l'AQQC ? Comment tenir compte des effets inattendus ou indirects des politiques publiques ?

P. Le Galès (CEVIPOF-CNRS) souligne d'abord le mérite de cette méthodologie qui, est une véritable stratégie de recherche et qui fournit une véritable alternative dans un contexte où l'on a tendance à réduire l'évaluation à une comparaison faite sur la base d'indicateurs quantitatifs. Il émet ensuite quelques réserves et critiques. Tout d'abord il remarque que cette méthode suppose en amont une stratégie de recherche collective de capitalisation des cas, stratégie qui reste rare pour l'instant.

Il pose ensuite la question de la rationalisation a posteriori des politiques publiques, démarche implicite à l'AQQC qui reste selon lui un pari intellectuel difficile, voire un vœu pieux, en tout cas une étape pas forcément nécessaire au regard des critiques qu'elle peut facilement susciter. Plutôt que de faire entrer à tout prix une politique dans le cadre d'analyse de l'AQQC en émettant de nombreuses réserves, pourquoi ne pas oser dire que telle politi-

que ne peut pas être prise en compte et essayer d'en analyser les raisons ? Pourquoi ne pas s'intéresser davantage aux cas qui ne marchent pas ?

Enfin il pose la question de l'unité de l'analyse. Contrairement à l'impression donnée par l'AQQC, les cas ne sont pas neutres, voire interchangeables. Comment l'auteur peut-il dès lors rendre les cas si « limpides », si linéaires ?

B. Rihoux répond à l'ensemble des remarques énoncées précédemment en déclarant qu'il comprend, pour en avoir été un également, les remarques de sociologues qualitativistes et rappelle que le pré-requis à l'utilisation de l'AQQC reste quand même une fine connaissance des cas. L'AQQC ne doit pas être utilisée de façon mécanique mais elle doit être un détour méthodologique pour revenir sur les cas, détour que C. Ragin appelle « le moment analytique ». Il récusé deux objections : celle selon laquelle la comparaison dans l'AQQC serait un but en soi et celle concernant le choix des cas. La comparaison reste un moyen tant que l'objectif de la recherche est l'identification de régularités entre les cas. Le choix des cas est pour lui un exemple significatif de tension possible entre les aspects politiques et techniques et il reconnaît qu'il n'a pas de réponse sur la manière de procéder. Concernant la prise en compte des effets inattendus, il remarque que des chercheurs ont déjà essayé d'intégrer ces effets et qu'il est nécessaire pour lui de simplifier pour comparer.

F. Varone complète la réponse précédente. Il considère que la prise en compte des effets inattendus se fait de toute manière en amont de l'AQQC et que celle-ci ne prend en compte que des politiques réalisées. Concernant la vision balistique des politiques publiques, il concède que l'AQQC puisse véhiculer une telle vision mais il remarque également qu'il est souvent possible de modéliser de façon simple le modèle causal à la base d'une politique publique. Par exemple, « si je donne des subventions, c'est pour augmenter... » Il prétend alors que si une analyse de cas fine permet d'arriver à un tel résultat, alors elle peut être intégrée à une AQQC.

Ne faut-il pas que l'AQQC rassemble un grand nombre de cas pour qu'elle soit significative ?

Pour B. Rihoux, il n'existe aucune règle en la matière et les AQQC actuellement pratiquées varient entre 3 et 4 628 cas ! Toutefois la plupart des recherches comprennent en général une dizaine de cas. Le nombre de variables doit être inférieur au nombre de cas pour que les situations ne se multiplient pas, ce qui implique d'avoir en amont une méthode ou une théorie permettant de sélectionner un certain nombre de variables. Enfin il conclut en précisant qu'il ne croit pas à l'existence d'une méthode miracle, que le pluralisme méthodologique est bénéfique pour la recherche et que toute bonne recherche repose avant tout sur une méthode adaptée à son objet. C'est dans cette optique que l'AQQC peut s'avérer utile.

TROISIÈME PARTIE

QUELLE PRISE EN COMPTE
DES INFORMATIONS
POUR QUELLE HIÉRARCHISATION
DES MONDES POSSIBLES ?

UNE TYPOLOGIE DES DÉMARCHES D'ÉVALUATION¹

Guillaume HOLLARD,
maître de conférences en sciences économiques
à l'Université de Marne-la-Vallée (OEP)

et Antoine VION,
maître de conférences en science politique
à l'Université de la Méditerranée (LEST)

La pratique de l'évaluation des politiques publiques tend à se diffuser sous les formes les plus diverses. Il suffit, pour s'en rendre compte, de comparer les revues spécialisées américaines, telles que *L'American Journal of Evaluation* ou *Practical Evaluation* ou encore les publications gouvernementales et intergouvernementales telles que les revues de l'OCDE ou de la Banque mondiale. Pour la France, les lectures d'*Économie et statistiques* ou d'*INSEE Premières* peuvent être utilement croisées avec celles de revues telles que *Politiques et management public* ou *Économie publique*. Les différents développements de l'évaluation sont désormais bien connus (Duran, Monnier, 1992 ; Kessler, Lascoumes, Setbon, Thoenig, 1998 ; Spenlehauer, 1999 ; Varone & Jacob, 2003, 2004a, 2004b, etc.). Les controverses récurrentes, mais rarement étalées au grand jour, qui traversent ce champ de l'évaluation des politiques publi-

¹ Une version préliminaire de ce texte a été présentée au 8^e congrès de l'AFSP à Lyon dans le cadre de l'atelier 20 « L'analyse des politiques publiques entre science des modèles et sciences de l'enquête ». Nous remercions les organisateurs de cet atelier (Pascale Laborier et Stéphane Nahrath), ainsi que les discutants (Jean-Claude Passeron et Jean-Louis Fabiani) et les autres participants. Pierre Lascoumes a bien voulu nous en faire un commentaire détaillé, les améliorations apportées lui doivent beaucoup. Nous remercions aussi Philippe Mossé, Francesca Petrella, et les participants au séminaire d'économie des conventions (Olivier Favereau, Franck Bessis, François Eymard-Duvernet, parmi d'autres) qui ont bien voulu se livrer à des discussions détaillées des versions successives.

ques, peuvent être vues comme une illustration supplémentaire des incompréhensions réciproques entre les tenants de la modélisation économique ou économétrique et les praticiens de l'enquête qualitative. Comme le soulignait récemment Patrice Duran, « l'évaluation a moins à craindre de son institutionnalisation proprement dite que de son enfermement dans des discours d'école » (Duran, 2002).

Notre propos n'est pas d'historiciser l'analyse des controverses entre écoles ou des pratiques des professionnels (Fourquet, 1976 ; Spenlehauer, 1999, 2003 ; Desrosières 2003a), ni de jeter un regard critique sur les postures des évaluateurs (Amiot, 1986 ; Mény, Thoenig, 1989 ; Fontaine, 1996 ; Spenlehauer, 1995, 1998, 2000 ; Thoenig, 2002). Nous défendons plutôt la thèse selon laquelle certaines controverses peuvent être mieux comprises si l'on explicite les manières de traiter l'information utilisée pour la démonstration.

Selon notre thèse, le traitement de l'information opéré dans le but de produire une expertise, et plus généralement un raisonnement en sciences sociales, peut être séparé en deux opérations logiques élémentaires : la prise en compte et la hiérarchisation. Nous nous inspirons sur cette question de la distinction établie par Latour (1999) entre prise en compte et ordonnancement à propos des modalités de la décision publique, mais nous proposons des définitions un peu différentes². Nous entendons ici par prise en compte la manière d'inclure des informations nouvelles dans le raisonnement. L'opération de prise en compte porte donc sur des informations. La hiérarchisation est une opération consistant à utiliser les informations prises en compte pour produire une hiérarchie des mondes possibles.

2 Le pouvoir de prise en compte doit permettre de déployer la « pluralité des mondes » grâce à des procédures, des « enquêtes pour détecter, rendre visible, et faire parler les différentes propositions candidates à l'existence ». Il suppose la mise en place de procédures de consultation pour constituer « le jury capable de juger des effets de chaque proposition sur les habitudes des autres ». Le pouvoir d'ordonnancement répond quant à lui au besoin d'unification, d'institution d'un monde commun. Il suppose une exigence de clôture (au moins temporaire) des controverses et des procédures contradictoires pour composer la hiérarchie optimale, c'est-à-dire finalement, décider.

Constatons que ces opérations sont banales et communes à l'ensemble des raisonnements mobilisés dans le cadre d'expertise. Un antagonisme entre ces deux opérations banales apparaît lorsqu'il convient d'explicitier le raisonnement suivi. Certains styles d'évaluation n'auront pas de mal à donner des gages de transparence sur les opérations de prise en compte mais laisseront dans l'ombre la manière dont le travail de hiérarchisation a été effectué. Réciproquement, d'autres styles d'évaluation seront à même d'explicitier complètement leur manière de hiérarchiser tandis que les opérations de prise en compte seront beaucoup moins vérifiables par un tiers. Avoir recours à un style d'évaluation impose un choix sur ce qui pourra être explicité. Les antagonismes et autres controverses qui traversent le champ de l'expertise pourraient avoir leur source dans une opposition logique qui tient à l'impossibilité de produire des raisonnements totalement transparents qui épouseraient les canons des sciences non humaines.

Nous allons exemplifier notre thèse en analysant en détail quatre grands styles d'évaluation, définis dans le tableau suivant : la démarche utilitariste (1), la démarche instrumentale (2), la démarche praxéologique (3) et la démarche procédurale (4). Nous montrerons, pour chacune comment procède le raisonnement afin de mettre en évidence la manière dont s'opèrent les opérations de prise en compte et de hiérarchisation.

La suite du document passe successivement en revue chacune de ces démarches, d'abord à un niveau général, puis en exemplifiant le raisonnement. En conclusion, nous étayerons notre thèse au regard des éléments rassemblés.

	Utilitariste	Instrumental	Praxéologique	Procédural
PRINCIPE	Analyse des préférences individuelles et des usages de biens	Réorganisation par les outils de gestion	Compréhension des pratiques administratives et des pratiques d'usagers	Irréductibilité des valeurs et primat de la délibération pour désamorcer les conflits et parvenir à un accord
STYLE DE DÉMARCHE ÉVALUATIVE	1. Constitution d'une base de données ou d'une liste de variable 2. Construction de	1. Classification d'activités individuelles ou collectives (nomenclature) 2. Valorisation des	1. Définition d'un ensemble de pratiques à analyser 2. Recueil de	1. Définition d'un enjeu ou d'un problème 2. Mise en place d'une procédure de

	l'enjeu et objectivation de la situation sur la base des données 3. Mise en concurrence d'explications causales par tests de variables 4. Hiérarchisation des mondes possibles 5. définition d'un optimum	activités par indexation 3. Comparaison des valeurs agrégées 4. Hiérarchisation des résultats	données sur la base d'hypothèses 3. Comparaison des explications causales issues de l'enquête 4. Hiérarchisation des explications sous la forme d'un cadrage théorique 5. Diagnostic des causes de succès ou d'échec	consultation 3. Prise en compte de l'ensemble des avis argumentés 4. Sélection des bons arguments 5. Synthèse des bons arguments
TYPES D'ENQUÊTES CONSTRUITES	- Recueil de données par questionnaires - Le cas échéant, enquêtes comportementales expérimentales ou en situation	- Collectes de données comptables - Création d'indicateurs - Recherche, instruction et sanction des fautes de gestion	- Enquêtes de fréquentation (statistiques descriptives) - Échelles d'attitudes - Enquêtes par entretiens - Observations directes - Recherche-action	- Collecte des arguments contradictoires - Recherche de preuves
ENJEUX DE DÉBAT	- Évolution des préférences - Représentativité des enquêtés - Existence d'une variable manquante	- Pertinence des indicateurs - Imputation comptable et imputabilité politique des dépenses publiques	- Épreuve de réalité des faits exemplifiés - Interprétation des observations	- Légitimité des dispositifs, des porte-parole, des arguments
PLACE DU POLITIQUE	Mise en œuvre des recommandations issues de l'expertise	Arbitrage budgétaire entre les catégories pré-définies	Interprétation et traduction dans le langage commun	Possibilité d'ignorer les résultats de l'expertise ; rivalité dans la mise en forme de la volonté populaire
PRISE EN COMPTE D'INFORMATIONS OU D'ACTEURS NOUVEAUX	Faible voire nulle sans modification de la procédure	Limitée aux catégories existantes	Bonne	Très bonne car pas de contrainte à priori
DEGRÉ DE HIÉRARCHISATION	Fort	Moyen-Fort	Moyen	Faible

1. LA DÉMARCHE UTILITARISTE

Lorsque nous employons ici l'adjectif « utilitariste », nous ne renvoyons pas seulement au courant de pensée utilitariste suivant une stricte approche historique des idées économiques. Nous envisageons plutôt ce terme comme le qualificatif d'une démarche générique centrée sur les comportements individuels (par opposition à l'observation de situations d'interaction). Les opérations d'évaluation qui peuvent être menées dans cette logique consistent à construire des modèles mathématiques, et à les soumettre à une série d'épreuves. Nous nous concentrerons sur la nature particulière du couplage opéré entre hiérarchisation et prise en compte dans ce type de démarche. Nous plaiderons l'existence d'une séquentialité qui initie le raisonnement en recensant toutes les informations à prendre en compte avant d'élaborer un raisonnement hiérarchisant ces informations. Nous commençons par décrire le raisonnement en termes généraux avant de l'exemplifier par un article récent d'analyse des politiques publiques, en l'occurrence un article de Thomas Piketty portant sur l'influence de la taille des classes sur la réussite scolaire.

Dans ce type de démarche, les éléments pris en compte sont limités et exprimés dans des catégories prédéfinies au sens où un sondage en question fermée contraint au préalable la nature de l'information recueillie. Les éléments se présentent à l'issue de ce travail sous forme d'une base de données ou d'une liste de variables. Une des caractéristiques de ce mode de raisonnement est de ne discuter l'adéquation entre la chose étudiée et sa mesure qu'à titre de préliminaire. L'analyse proprement dite est distincte de cette phase de discussion. Le travail de prise en compte, ainsi que la discussion critique inhérente à celui-ci, doivent être clos avant de passer à l'analyse proprement dite. Les informations sont listées mais une seconde phase, moins visible, consiste à réaliser un appariement entre les données ou variables utilisées et les notions manipulées. Par exemple une analyse comparée des systèmes productifs de différents pays s'appuie sur la comparaison de données particulières : les niveaux de PIB. Le calcul du PIB lui-même se fonde sur une longue série de traductions et classifications des transactions marchandes. Les insuffisances du PIB sont largement soulignées et abondamment commentées : le débat occupe

une large part des programmes de première année d'économie et figure en bonne place dans les manuels élémentaires de cette discipline. À titre d'exemple, on souligne souvent que les transactions qui franchissent la frontière entre le marchand et le non-marchand altèrent la mesure du PIB. Si monsieur épouse sa femme de ménage (ou madame son chauffeur), le PIB diminue. Que dire par ailleurs d'un pays où le troc est largement répandu et la production agricole largement auto consommée ? On conçoit bien que résumer en un seul chiffre l'ensemble du système économique national suppose quelques approximations. Quand il s'agit ensuite d'utiliser les données ou les variables retenues, la discussion est close. Par convention, les termes « PIB » et « richesse produite » sont devenus équivalents. On prête à Alfred Binet, l'inventeur du premier test de QI cette phrase emblématique : « L'intelligence ? C'est ce que mesure mon test. » On pourrait dire de manière similaire : « La richesse ? C'est ce que mesure le PIB. » On instaure donc une correspondance biunivoque entre une notion abstraite (l'intelligence, la richesse) et sa mesure (le QI, le PIB). La phase de prise en compte légitime, ou du moins tente de rendre légitime, un usage des données comme si celles-ci suffisaient à capturer l'essentiel du contexte duquel elles sont extraites.

Une fois close l'opération de prise en compte, il est possible de définir l'état du monde présent, sous la forme d'écart de richesse entre les pays, d'inégalités des niveaux d'intelligence, etc. Le travail du modélisateur consiste alors à élaborer un modèle explicatif qui permet de définir les modalités d'intervention sur ce monde présent pour produire un éventail de mondes possibles. Ces mondes possibles sont ensuite hiérarchisés au regard de critères explicites (critère rawlsien, critère égalitariste, critère d'efficacité, etc.). Le modélisateur se place dans une position de neutralité par rapport aux valeurs puisque l'opération de hiérarchisation se fait au nom de critères explicites et extérieurs au modélisateur. C'est la partie noble de son raisonnement, là où s'exprime pleinement le talent de celui qui conduit l'expertise. Le chemin qui conduit de la prise en compte initiale aux conclusions du raisonnement est clairement balisé et, idéalement, ne laisse rien dans l'ombre. Il est possible de produire un raisonnement transparent au sens où celui-ci peut-être conduit au nom de la logique seule. Dans cette seconde étape, le raisonnement épouse les

canons de la science : il est logique, souvent formel et peut être reproduit à l'identique par un tiers. Il donne donc des garanties de transparence qui, associées au caractère quantitatif de l'analyse, rendent l'expertise difficilement discutable.

Ce style d'évaluation produit des résultats de la forme « *si X alors Y* ». Typiquement, « si les objectifs de la décision sont ceux-ci (réduire le chômage, diminuer les inégalités, etc.) alors il convient de mettre en œuvre telle politique, mettre en œuvre un nouveau type de contrat de travail, accroître tel type de financement, etc. ». Les destinataires de l'évaluation n'ont d'autre choix que d'appliquer, ou non, les recommandations résultant de l'évaluation. Charge à eux de les rendre légitimes, ou au moins acceptables.

Venons en à notre exemple, concernant l'impact des tailles de classes sur la réussite scolaire. Thomas Piketty présente dans cette étude³ les résultats d'une analyse économétrique à partir d'une base de données collectée par la Direction des études et de la prospective du ministère de l'Éducation nationale. Nous avons retenu cet article pour exemplifier notre propos général du fait qu'il s'agit d'un travail largement salué comme ayant renouvelé la vision de ce sujet classique. Il est considéré comme une bonne évaluation. L'annonce du plan résume les grands traits de la démarche :

« La suite de cette étude est organisée de la façon suivante. La section 2 décrit les principales caractéristiques techniques des données utilisées. La section 3 plante le décor en présentant un certain nombre de statistiques descriptives concernant la taille des classes et les inégalités de réussite scolaire dans les écoles primaires françaises. La section 4 présente les résultats de nos estimations de l'impact de la taille de la classe sur la réussite scolaire, en commençant par les estimations « naïves », puis en développant la méthode fondée sur les discontinuités liées au franchissement des seuils d'ouverture des classes. Nous utilisons ensuite les résultats de ces estimations dans la section 5 pour évaluer dans quelle mesure

³ Nous nous basons ici sur un document de travail accessible en ligne : <http://www.jourdan.ens.fr/piketty/fichiers/public/Piketty2004b.pdf>

des politiques de ciblage en faveur des écoles défavorisées plus ambitieuses que celles menées aujourd'hui pourraient permettre de réduire les inégalités scolaires. Dans la section 6, nous présentons l'impact de la ségrégation sociale sur la réussite scolaire, afin notamment de comparer l'ampleur de ces effets à ceux des tailles de classes réduites. Enfin, la section 7 présente un certain nombre d'extensions et de questions ouvertes. »

La partie 2 de l'article décrit les informations contenues dans la base de données et la partie 3 ajoute d'autres informations à prendre en compte. Il est important de noter que l'apport d'informations nouvelles est limité à ces deux sections du texte. L'appariement entre notions (la réussite scolaire, les élèves défavorisés, etc.) et mesures est fixé à ce niveau. Par la suite la réussite scolaire sera assimilée ce que mesurent les tests de la DEP, la notion d'élève défavorisé se confondra avec la PCS des parents, etc. De plus, les marqueurs du contexte disparaissent. On ne parle plus d'un échantillon d'école à une date donnée mais du système scolaire français en général.

Ces opérations d'appariements et de décontextualisation sont caractéristiques des opérations de prise en compte dans ce style de démarche. Le point important pour notre propos est de noter que ces opérations recèlent une part d'ombre qui, quelque soit la bonne foi de l'évaluateur, ne peut être rendue complètement transparente. On s'en remet nécessairement à la fiabilité supposée des institutions qui ont produit les données, au prestige et à la renommée de l'auteur.

Une fois ce préalable franchi, le raisonnement devient transparent (l'auteur rend même les données accessibles sur son site Internet). Chacun peut donc combiner les données comme il lui convient. L'auteur utilise un modèle économétrique simple. L'apport de l'analyse est d'utiliser l'effet de seuil introduit par l'existence d'un seuil d'ouverture ou de fermeture de nouvelles classes. Ainsi, il peut observer dans un environnement social constant, l'effet d'une réduction ou d'une hausse du nombre d'élèves par classe. Reproduire les mêmes calculs que l'auteur conduira aux mêmes résultats. La section 5 est emblématique de la démarche décrite : le modèle est étendu à des situations hypothétiques (ce qui se passerait si l'on réduisait la taille des classes). Ces

situations sont comparées entre elles au regard d'un critère rawlsien qui se concentre sur la situation des plus défavorisés (une situation est d'autant meilleure que la situation des plus défavorisés est meilleure). L'optimum rawlsien est calculé de manière formelle. L'auteur en tire des recommandations pratiques et chiffrées : une réduction de 40 % des inégalités constatées peut-être atteinte, à moyens constants, en diminuant la taille des classe en ZEP jusqu'à 18 élèves (contre 22 actuellement) au prix d'une légère remontée de la taille des classes hors ZEP, qui passeraient de 23.3 à 24.2 élèves. Le message final est bien de la forme « si vous souhaitez réduire les inégalités scolaires, alors voici ce qu'il convient de faire ». On peut difficilement être plus prescriptif. La section 6 expose un modèle théorique prenant en compte simultanément deux sources d'inégalités liées à la taille des classes et la ségrégation scolaire. La section 7 boucle le raisonnement en pointant les informations manquantes qui permettraient d'étendre la conclusion.

Les critiques adressées à style d'expertise sont connues, et presque imparables. Au regard de ce qui précède, il est possible d'anticiper le fait qu'elles portent principalement sur la phase de prise en compte puisque que c'est à ce niveau qu'interviennent les (inévitables) bricolages méthodologiques. L'exemple le plus courant concerne l'existence d'une causalité non prise en compte. Imaginons par exemple que l'enseignement en ZEP soit plus axé sur une pédagogie participative qui vise le développement harmonieux des enfants plutôt que leur capacité à réussir des tests formels. L'outil de mesure de la réussite scolaire retenu (des tests standardisés) trouverait donc sans surprise un meilleur résultat hors ZEP et ne ferait en fait que constater un écart de méthode pédagogique, et non pas un écart de réussite scolaire. Dans cet exemple (fictif), l'adéquation entre les données exploitées et la notion de réussite scolaire peut également être mise en doute puisque la réussite scolaire peut ne pas prendre le même sens en ZEP et hors ZEP. La méthode d'expertise employée ne permet pas de se prémunir contre des critiques de ce type, alors qu'il est possible de reproduire les calculs qui ont mené aux conclusions.

Le principe d'une évaluation portée par un raisonnement quantitatif a une longue histoire en France et remonte au moins aux ingénieurs économistes tels que Cournot ou Dupuit. Cette tradition de serviteur de l'État soucieux de

développer les outils d'une véritable science économique s'est prolongée dans la création en 1930 de l'influente société d'économétrie. Il en a résulté un recours privilégié au raisonnement hypothético-déductif qui a durablement marqué le champ de l'évaluation. L'analyse *ex-post*, qui inclut l'exemple traité ci-dessus, trouve sa pleine justification dans l'analyse d'expérience naturelle. On regroupe sous ce terme des situations particulières où une partie de l'échantillon a été soumise à une transformation sociale et l'autre non. Par exemple, on peut comparer deux cohortes de chômeurs qui ont été aléatoirement affectées à deux dispositifs d'aide à l'emploi différents (Cahuc, Zylberberg 2004) ou comparer les bénéficiaires et les non bénéficiaires de dispositifs de formations de chômeurs pour en mesurer les impacts (Van der Linden, 2001). De ce fait, les variables manquantes peuvent être supposées égales d'un groupe à l'autre, ce qui rend l'opération de prise en compte moins vulnérable. Les expériences naturelles fournissent un point d'appui privilégié puisque le raisonnement devient difficile à remettre en cause. La place laissée au politique est par conséquent restreinte à la mise en œuvre des recommandations de l'expertise.

Cependant, l'analyse peut également opérer *ex-ante* et l'arsenal mis en place n'est pas moins impressionnant dans ce cas. Le calcul économique, abondamment employé en matière de transport, d'environnement et, plus récemment, de santé, a élaboré en effet un ensemble de dispositifs dont l'objectif affiché est de pouvoir rendre comparables *ex-ante* les différentes dimensions d'une décision publique, pour pouvoir *in fine* hiérarchiser les différents projets. La hiérarchisation opère en rabattant chacune des dimensions de la décision sur un axe unique, gradué en unité monétaire. Un haut degré de sophistication est atteint pour fournir des équivalents monétaires à des aspects aussi peu monétisables a priori que la qualité de vie (Geniaux et al. 2001) ou les effets locaux de la pollution atmosphérique (Nicolas et al, 2002). Les développements récents vont jusqu'à enquêter sur les préférences futures des agents pour donner une valeur à des politiques à mettre en œuvre. L'évaluation contingente est emblématique de cette approche. Elle marque une évolution vers la prise en compte de données issues d'interviews, alors que l'analyse traditionnelle reposait sur la prise en compte d'actions observées ou de quan-

tités objectives. En dépit de ses défauts (Flachaire et Hollard, 2006), la méthode d'évaluation contingente connaît un développement rapide à travers le monde.

S'agissant d'anticiper des pratiques futures, l'évaluation *ex-ante* se base sur des informations moins solidement établies que dans le cas d'une expérience naturelle. De ce fait, la place accordée au politique par les experts diffère quelque peu et semble évoluer vers une reconnaissance du rôle des élus dans la détermination des choix futurs (Crozet 2004). À titre d'exemple, le « coût du mort » qui permet de déterminer l'opportunité de nouveaux investissements en matière de sécurité, conduisait à disqualifier les choix de certains élus qui semblaient sur-investir dans des équipements de sécurité (« zone 30 », ralentisseurs, etc.). Or, quelques années plus tard, le coût du mort ayant été réévalué, ces mêmes équipements étaient devenus rentables au regard des nouveaux critères en vigueur.

2. LA DÉMARCHE INSTRUMENTALE

Nous entendons par démarche instrumentale d'évaluation toute entreprise visant à construire un dispositif technique produisant des valeurs sur la base d'outils de gestion⁴ destinés à optimiser la gestion des moyens d'une organisation. Ce type de dispositif repose donc sur la réalisation et la routinisation de procédures automatiques de catégorisation et de comptage qui formatent a priori la hiérarchie des informations pertinentes pour conférer sa valeur à l'activité observée. Ce style de démarche évaluative se caractérise donc par le fait que des classifications sont établies a priori sous forme de nomenclatures, chaque item étant ensuite valorisé par indexation (voir encadré pour l'exemple qui suit), puis les valeurs agrégées comparées et les agrégats hiérarchisés.

⁴ On appellera outil de gestion « un ensemble d'outils formalisés (textes, logiciels...) destinés à coordonner des opérations de gestion (du travail, des produits, des processus, des personnels...) au moyen de procédures considérées comme efficaces » (Boussard, Mercier, Tripier, 2004).

Trois points de discussion nous semblent importants pour comprendre ce qui est en jeu dans ce type de démarche : le travail de construction des modes de sélection des informations, le caractère difficilement réversible de l'instrumentation, et les effets de standardisation des comportements qu'elle engendre.

Le travail de construction des modes de sélection des informations est toujours, dans ce type de démarche, l'objet de vifs débats entre les techniciens chargés de configurer l'outil, les autorités politiques et administratives commanditaires, et les services dont la gestion sera affectée par l'outil. Pour ne prendre qu'un exemple récent, la mise en œuvre de la LOLF⁵ dans les administrations nationales françaises met au centre des discussions la pertinence des informations susceptibles de servir à des opérations de mesure de la performance⁶. Faut-il, par exemple, pour évaluer l'activité d'une équipe d'enseignants d'une UFR, définir le pourcentage du service individuel effectué en recherche, en enseignement, et en administration, et faire une moyenne par UFR, par statuts, etc. ? Ce mode de mesure pose le double problème de la catégorisation des tâches (ne prépare-t-on pas un cours lorsqu'on cherche et réciproquement ?) et de la mise en équivalence des statuts : doit-on s'en tenir aux grades et échelons ou faut-il introduire des variables plus fines comme le taux d'encadrement par établissement, voire par formation, ce qui nuancerait très fortement le postulat selon lequel la charge d'enseignement d'un maître de conférences enseignant par exemple dans un diplôme de Licence de psychologie équivaldrait à celle d'un maître de conférences en mathématiques. Ces questions ne sont évidemment pas détachées des anticipations que font les protagonistes des usages de l'instrument en termes de mesure des performances⁷.

5 Pour une présentation technique approfondie de la LOLF, voir le numéro spécial de la *Revue française de finances publiques* de 1999.

6 Pour une discussion de la mise en œuvre de la LOLF dans la magistrature, cf. Pallez, 2002.

7 Alain Desrosières (1992) montre ainsi, par exemple, que le travail des statisticiens est perpétuellement traversé par la tension entre le souci d'optimiser la mesure et la peur panique de développements potentiels d'usages politiques non maîtrisables. La sociologie des sciences inspirée

Si le travail de construction de l'instrument d'évaluation n'est pas neutre, il n'est pas non plus complètement libre, dans la mesure où il est conditionné par la résolution de contraintes techniques parfois difficilement surmontables. Que se passe-t-il, dans notre exemple précédent, si le travail de programmation informatique permet la prise en compte de la catégorie « taux d'encadrement par diplôme », mais que le travail de modélisation économétrique destiné à la comparaison des performances ne permet pas de prendre simultanément en compte l'ensemble des taux par diplôme pour mesurer la charge de travail effective d'un enseignant et d'agrèger ensuite ces charges par établissement ? Certains ont à juste titre insisté sur le fait que la sélection des informations opère une réduction de la complexité organisationnelle. Mais il faut aussi souligner, en écho aux débats de sociologie des sciences sur le statut ontologique des instruments, que les limites techniques éprouvées dans la construction du dispositif affectent le sens même de l'activité évaluée. S'il est donc toujours possible de produire des réinterprétations par la production de nouvelles convergences et mises en relation des informations (Akrich, Callon, Latour, 1988), cela ne met qu'exceptionnellement en cause la légitimité de l'instrument lui-même, qui tend à s'imposer progressivement à l'ensemble de la communauté intéressée (Desrosières, 1993 ; Bagla-Gökalp, 1996), à moins d'arbitrages politiques décidant le recours à des formes d'instrumentation alternatives exclues par le passé (Barthe, 2003).

Tout un pan de la recherche sur l'action publique s'interroge aujourd'hui sur ce phénomène de « durcissement » progressif des instruments d'évaluation⁸. L'exemple de l'évaluation de la gestion hospitalière en France est sans doute le plus significatif en la matière. La Direction Générale de la Santé a introduit au début des années 1980 dans la gestion hospitalière un instrument d'analyse

par Simondon a montré que la construction d'instruments, loin d'être une technicité neutre, met constamment en jeu l'arbitrage entre des voies de développement différentes.

⁸ On peut utilement faire un point sur cette question en se reportant à l'introduction de Pierre Lascoumes et Patrick Le Galès, et aux contributions de Philippe Bézès et Patrick Le Galès dans le livre *Gouverner par les instruments*, paru l'an dernier aux Presses de Sciences PO (Lascoumes, Le Galès, 2004).

des séjours d'hospitalisation destiné à évaluer l'activité des établissements en comparant les types de pathologie et les types de durée de séjour des patients. Cet instrument, le Programme de Médicalisation des Systèmes d'Information (PMSI), était inspiré de travaux américains réalisés le Professeur Fetter et son équipe à l'Université de Yale, consistant à recouper systématiquement des informations sur les malades et leur traitement en *diagnostic related groups* (DRG). Dans cette logique, le PMSI a défini, pour les activités dites « de court séjour » (médecine, chirurgie et obstétrique)⁹ des opérations de recouplement des informations médicales (identification des types de pathologie et de traitement en groupes homogènes de malades – GHM –) et des informations administratives (détermination de la longueur du séjour, calcul du prix de journée) permettant aux établissements hospitaliers de rationaliser la gestion des lits, et de comparer leurs performances. Au moment du lancement officiel de cette réforme, au milieu des années 1980, le Directeur des hôpitaux présentait le PMSI, à l'issue de nombreux débats, comme une simple amélioration du système d'information sur les malades permettant un contrôle interne de la gestion des lits. Si un instrument comme le PMSI s'est progressivement imposé, c'est parce qu'il offrait aux médecins des hôpitaux les moyens de se positionner par rapport aux cliniques et de comparer entre eux les effets de leurs techniques de soin sur la gestion des lits. C'est parce que l'instrument d'évaluation est aussi un instrument de valorisation d'une communauté professionnelle qu'il tend à s'imposer. Comme le suggère Giddens (2000), la modernisation n'est pas seulement un processus de rationalisation technique dominant, c'est aussi un processus qui développe la réflexivité des individus sur leurs pratiques.

La démarche d'évaluation a ensuite été précisée en 1991 par une disposition réglementaire obligeant les hôpitaux à constituer pour chaque patient un résumé de sortie standardisé (RSS) recoupant un diagnostic principal, des diagnostics associés, des actes techniques, l'âge, le sexe, etc. L'introduction d'un logiciel groupeur a ainsi permis d'opérer une nomenclature de 540 produits à partir de laquelle les prestations hospitalières devenaient comparables.

⁹ La tendance actuelle est à l'extension des domaines d'activité couverts par la PMSI.

C'est la construction de cette comparabilité des données qui a ensuite permis aux économistes de la santé, et particulièrement les économètres, de concevoir des modèles d'agrégation permettant l'indexation de l'activité des établissements sur un standard de performance. La construction de ce nouveau savoir a ensuite permis à l'administration des hôpitaux, dans le cours d'une réforme de décentralisation créant les agences régionales d'hospitalisation, d'opérer une valorisation économique de l'activité. Cette valorisation a pris la forme suivante : « la valorisation économique relative des GHM est ensuite construite à partir d'un échantillon d'établissements participant à ce qu'il est convenu d'appeler "l'étude nationale des coûts" ». Ces établissements publics ou participant au service public disposent tous (Charretton et Fischesser, 1995) d'une comptabilité analytique détaillée rendant possible le calcul du coût complet de chaque séjour¹⁰. Lorsque la base des coûts des séjours est constituée, il ne reste plus qu'à agréger les dépenses au niveau de chaque GHM. Enfin, le coût d'un GHM est exprimé sur une base relative, en nombre de points ISA (Indice synthétique d'activité). Les établissements hospitaliers peuvent alors décrire la lourdeur économique et médicale des patients traités dans leur structure et évaluer leur niveau d'activité en valorisant leur case-mix par l'échelle de points ISA par GHM. » (Lenay, 2000).

¹⁰ Depuis 1995, le ministère en charge de la Santé publie chaque année un BO présentant la méthodologie et les résultats de l'étude nationale des coûts par GHM. On trouve en particulier dans le fascicule tous les détails concernant la comptabilité analytique mise en œuvre, les unités d'œuvre permettant la ventilation des dépenses liées aux actes médico-techniques (Indices de coût relatif) et les méthodes de ventilation des coûts indirects (note de l'auteur cité).

Valorisation en point ISA

document de l'agence technique de l'information hospitalière¹¹

Chaque année dans MAHOS est intégrée une liste de lettres-clés permettant de valoriser en points ISA l'activité externe des établissements.

Cette activité est déclarée par les établissements par le biais de fichiers produits par le logiciel FICHSUP.

Depuis 2001, les établissements doivent fournir deux fichiers correspondants à cette activité externe :

- fichier concernant l'activité externe hors urgence,
- un fichier concernant l'activité des urgences.

Ces deux fichiers contiennent les mêmes informations et sont donc traités de la même manière pour aboutir à un nombre de points ISA. Le nombre de points correspondant à l'activité externe correspond à la somme des points ISA calculés sur ces deux fichiers.

Avant 2001, un seul fichier était fourni.

Ces fichiers fournissent pour chaque lettre-clé la somme des coefficients facturés.

La valorisation en point ISA est donc obtenue en faisant la somme des produits pour chaque lettre clé du nombre de coefficients par le nombre points ISA correspondant à un coefficient lu dans la table de l'année voulue.

Le nombre de points ISA correspondant à un coefficient est obtenu, pour la totalité des lettres clés « Officielles » (publiée au Journal officiel), par la division du coût en centimes de la valeur officielle de lettre clé, par la valeur du point ISA en centimes dans l'Étude nationale de coût (coefficient de conversion).

Pour les activités de scanner et d'IRM la valorisation s'effectue de manière plus indirecte.

En effet, pour ces 2 types d'équipements, la valorisation ne peut pas s'effectuer directement à partir de la valeur d'une lettre clé, mais elle s'effectue à partir de la valeur d'un forfait qui dépend du type de machine ainsi que de son ancienneté ; ces valeurs de forfaits sont également publiées au JO. De plus, il existe pour chaque équipement une valeur de forfait normal et une valeur de forfait réduit, correspondant à la prise en compte de l'amortissement de l'équipement.

Pour effectuer cette valorisation nous avons créé des lettres clés particulières permettant de coder la valeur du forfait normal et réduit, ainsi que le nombre d'examen réalisés à la valeur de ces forfaits.

11 Nous reproduisons ici à titre illustratif l'un des modes d'indexation relatif aux activités externes des hôpitaux, le MAHOS, mais pour une vision exhaustive de cette activité, il faudrait aussi reproduire d'autres modes de valorisation comme le MIGAC relatif à certaines activités internes.

Ces lettres clés sont :

Pour le scanner :

Valeur du forfait normal : SCFN_x

Valeur du forfait réduit : SCFR_x

Nombre de scanner réalisés à forfait normal : SCNN_x

Nombre de scanner réalisés à forfait réduit : SCNR_x

Pour les IRM :

Valeur du forfait normal : RMFN_x

Valeur du forfait réduit : RMFR_x

Nombre d'IRM réalisées à forfait normal : RMNN_x

Nombre d'IRM réalisées à forfait réduit : RMNR_x

Le 'x' correspond à un identifiant permettant de saisir l'activité de plusieurs équipements de scanner ou d'IRM.

Pour la valorisation en point ISA, le calcul est le suivant :

On multiplie la valeur du forfait normal de l'équipement n° 1, par le nombre d'exams réalisés sur cet équipement à la valeur de ce forfait que l'on divise, comme pour les lettres clés standard, par la valeur du point ISA en centimes dans l'Étude nationale de coût (VCENC).

Par exemple pour un scanner portant le numéro 1 on obtiendrait pour ce qui est de son activité à forfait normal :

Valorisation en ISA de l'activité à forfait normal = $SCFN_1 \times 100 \times SCNN_1 / VCENC$

(si SCFN₁ est en centimes)

Ce qui est fourni dans la table que vous pouvez consulter correspond au facteur 1/VCENC. , pour les lettres clés SCFN_x, SCFR_x, RMFN_x, RMFR_x.

Pour récupérer une table en format texte utiliser la commande d'enregistrement de votre navigateur pendant la visualisation de la table en format texte.

Dans une enquête relative à l'introduction de l'outil de gestion ISA dans le milieu hospitalier, Jean-Claude Moisdon (2000) a montré comment le point ISA devenait progressivement l'étalon de référence de la capacité de gestion des directeurs d'hôpitaux, au point qu'une donnée de base des échanges professionnels devenait la réponse à la question : « Quelle est la valeur de ton point ISA ? » Passé en quelques années d'un simple instrument de connaissance des malades et de gestion des lits à un véritable instrument d'évaluation

de la performance des hôpitaux, le PMSI a ainsi connu une série de développements plus ou moins inattendus impliquant des réorientations significatives de l'action gouvernementale. En effet, les agences régionales, sur la bases de modèles d'incitation, taxent aujourd'hui les établissements ayant un point ISA élevé, et restituent des moyens aux établissements sous-dotés. Bel exemple d'évolution des usages d'un instrument d'évaluation dans le sens d'un durcissement de la contrainte financière.

Les critiques adressées à cette démarche d'évaluation sont de deux ordres. Tout d'abord, l'établissement de valeurs de référence permettant l'indexation provoque une décontextualisation qui peut aboutir à des absurdités pratiques. Pour ce qui concerne l'évaluation de la performance des services ambulanciers en Grande-Bretagne, Chris Hood (2004) a bien montré les effets de standardisation des comportements provoqués par ce type d'indexation. Dans le cas qu'il étudiait, l'étalonnage sur une valeur moyenne de 8 minutes de trajet dans un périmètre donné avait eu pour conséquence que les ambulanciers, quelle que soit la distance du trajet, mettaient un point d'honneur à mettre 8 minutes, prenant leur temps pour les trajets courts, ou déployant un zèle énorme pour les trajets longs.

Dans un autre domaine d'activité, il peut ressortir que les critères d'évaluation de la qualité paysagère (variété des couleurs, lignes de fuites, etc.) s'appliquent très bien à une décharge de vieilles voitures (Charretton et Fischesser, 1995). On touche ici à une limite de la démarche instrumentale d'évaluation. Le pouvoir de conformation des activités sociales aux *optima* définis par l'instrument n'est jamais assuré, dans la mesure où les catégories prédéfinies subsument une série de traits qui peuvent être exemplifiés par une multitude d'objets ou de comportements.

Par ailleurs, il ne faut pas négliger, au delà du jeu des catégories, la capacité des individus à jouer avec les catégories, et à développer aussi ce que Michel de Certeau (1971) appelait des « pratiques rusées ». Pour ce qui concerne notre exemple de gestion des hôpitaux, le durcissement de l'indexation de l'allocation des ressources sur les indices affectés aux DRG aux États-Unis a eu des effets inattendus : les groupes professionnels visés ont configuré et diffusé

des logiciels de gestion destinés à « optimiser le patient » au regard des règles de dotation : pour une série de symptômes donnés, le logiciel choisit la catégorie de pathologie la plus rentable.

Tous ces éléments permettent de comprendre les critiques les plus régulièrement adressées à cette démarche d'évaluation. Elle opère à l'évidence une réduction de la complexité des organisations ou des mondes sociaux. Le travail de sélection des informations pertinentes permet évidemment de hiérarchiser les priorités d'action, mais il affecte durablement le sens de l'activité concernée. L'absurdité de certains résultats, comme la « très grande qualité paysagère » d'une décharge de voitures peut ainsi conduire à mettre en doute l'efficacité de cette démarche d'évaluation. Ces critiques, si elles sont valides, ne suffisent pourtant pas à disqualifier la démarche, sauf à s'en tenir à une posture anti-institutionnelle. On ne peut en effet occulter le fait que la critique en elle-même offre des alternatives et permet de reconfigurer les outils d'évaluation. De plus, l'intérêt de cette démarche, dans une perspective démocratique, tient dans le fait que les arbitrages politiques entre des voies de développement différentes sont nombreux. Hiérarchiser les mondes possibles et arbitrer les méthodes, même si cela impose une certaine technicité, ne relève pas forcément d'une logique de contrôle de type disciplinaire. La critique foucauldienne porte lorsqu'elle s'applique aux usages de cette démarche, mais elle s'applique difficilement aux opérations cognitives sur lesquelles repose la logique même de cette démarche. Ce qui pose problème en effet, c'est l'inclination inconsidérée au durcissement des systèmes d'indexation et le pilotage centralisé des outils de gestion.

3. LA DÉMARCHE PRAXÉOLOGIQUE

La notion de démarche praxéologique d'évaluation s'applique, dans l'acception que nous lui donnons, à tout travail de diagnostic reposant sur la compréhension des pratiques d'agents ou d'usagers et/ou les interactions entre eux. Plusieurs méthodes peuvent être utilisées, de la statistique descriptive de type enquête sur les publics à l'observation participante (Weller, 2001 ;

Dubois, 2001 ; Boussard, Mercier, Tripier, 2004). Quelle que soit la méthode employée, la démarche consiste d'abord à définir un ensemble de pratiques à analyser, à dresser ensuite un nombre plus ou moins important d'hypothèses, à recueillir des données sur cette base, puis à comparer des explications causales issues de l'enquête, et enfin à hiérarchiser ces explications sous la forme d'un cadrage théorique permettant de poser le diagnostic des causes de succès ou d'échec d'une politique. La prise en compte d'informations diverses sur les pratiques peut permettre de construire des « exemplaires¹² » qui révèlent des phénomènes que les hypothèses ne pouvaient prévoir et qu'il aurait été impossible d'identifier sans procéder au travail d'enquête (Passeron, 1998).

Un nombre considérable d'évaluations est effectué chaque année sur ces bases. Nous ne retiendrons qu'un exemple, qui nous paraît assez significatif de l'intérêt de cette démarche. Il concerne une enquête menée par Pierre Tripier et ses collaboratrices à propos de la mise en place d'une politique de tri sélectif dans une ville de la banlieue parisienne. Afin de comprendre les difficultés de mise en œuvre de cette politique, l'équipe de sociologues a suivi la tournée des rippers et enquêté sur le parcours des bacs de déchets. L'un des principaux résultats des observations et des entretiens réalisés avec les personnels a été le constat selon lequel les employés de l'entreprise de collecte à laquelle cette gestion avait été déléguée par la ville changeaient parfois leurs horaires de tournée sans prévenir les gardiens d'immeubles qui avaient pour mission d'acheminer les bacs. Les gardiens se trouvaient donc exclus du cercle de tri, et traînaient les pieds dans la mise en œuvre du dispositif. La principale préconisation du rapport remis à la ville commanditaire a été de réintégrer les gardiens dans ce cercle par des processus de motivation de type réunions, pots de fin d'année, etc.

12 Un exemplaire est « un cas probant qui procure un accès épistémique aux traits qu'il exemplifie. Un cas probant « révèle, dévoile, transmet des aspects de lui-même » (Elgin, 1992). Il peut ainsi revivifier l'évident (donné par le sens commun ou un corpus de textes), ou au contraire exemplifier des traits excessivement inapparents au regard des hypothèses d'arrière-plan. Sur l'usage de la notion d'exemplaire pour comprendre l'opération de construction typologique (Vion, 2004).

Le plan du compte-rendu de cette enquête va nous permettre de progresser dans notre réflexion :

« Nous verrons dans le cas présenté ci-après comment les résultats décevants des opérations de tri sélectif dans les zones d'immeubles peuvent être analysés comme la rencontre, ratée, entre les définitions des organisateurs du tri et celles des gardiens d'immeubles. Et nous verrons surtout que le malentendu se fait sur la place occupée par les gardiens, sur leur appartenance bafouée au "cercle de tri". Ce diagnostic du dysfonctionnement montre que les explications de premier niveau formulées par les acteurs de l'organisation ne sont pas justifiées : l'échec relatif du tri sélectif n'est pas dû, comme le laissent entendre certains, à une mauvaise communication des objectifs et des méthodes aux habitants. Un renforcement du dispositif d'information auprès de la population, pour renforcer les apprentissages et les changements des habitudes, ne serait pas la solution la plus pertinente. Nous verrons également que les diagnostics sociologiques classiques (analyse des relations de pouvoir, analyses de type culturel) ne permettent pas de remonter aux causes premières, causes que nous faisons apparaître à l'aide d'un autre modèle théorique. » (Boussard, Mercier, Tripiet, 2004)

À la lecture de ce plan, nous voyons d'abord que la démarche praxéologique est envisagée comme un mode de réponse à des échecs difficilement compréhensibles, où la rationalité gestionnaire est mise en défaut. Le concept de « résultats décevants » a bien des affinités avec la démarche utilitariste, mais l'analyse des résultats en question n'est pas outillée. Ce n'est pas le problème. Un économiste comme Piketty aurait essayé de calculer les rendements par site pour identifier les points précis où la collecte a un rendement inférieur aux attentes (en fonction de variables démographiques et autres), puis aurait commandé un questionnaire aux habitants pour tester des variables afin d'optimiser le tri (surcroît d'information, politique d'incitation financière ou autres). Notre équipe part au contraire d'un constat du commanditaire pour définir son objet d'enquête. La seconde phrase permet de comprendre ce qui justifie ce type de démarche. Trivialement : les dysfonctionnements ne sont pas compréhensibles si l'on ne regarde pas de près ce qui se passe et qu'on

n'écoute pas les gens. C'est donc l'enquête qui en révèle la nature. Il n'y a pas de schématisation pertinente a priori ! Puisque ce sont justement les catégories ou les pratiques sociales tacites qui posent problème, l'objet de l'enquête sera de les reformuler. Le social est affaire de prise en compte. Le terrain prime, il n'y a pas lieu de procéder par le haut et de hiérarchiser hors contexte. Lorsque Boussard, Tripiet et Mercier disent que « un renforcement du dispositif d'information auprès de la population, pour renforcer les apprentissages et les changements des habitudes, ne serait pas la solution la plus pertinente », il est intéressant de noter qu'ils vont réfuter la démarche instrumentale par un argument instrumental : maintenant qu'on sait cela, on ne va quand même pas investir du temps et de l'argent dans l'instrumentation. On va plutôt traiter humainement les gardiens. En soi, la solution n'est pas moins rationnelle, dans la mesure où elle peut avoir un impact non négligeable sur le processus. Mais elle contraste avec les démarches précédentes dans la mesure où elle n'a pas pour objet de mettre en place les outils permettant de mieux mesurer les rendements ou de tester a priori l'efficacité de la variable « gardien d'immeubles ».

Cette démarche d'évaluation est donc, comme les précédentes, beaucoup plus un style de raisonnement qu'une simple méthode d'enquête. Il serait en effet réducteur de prétendre que la démarche praxéologique se caractérise par l'usage de méthodes qualitatives comme l'observation directe, les entretiens, *focus groups*, et autres. De nombreux usages de la statistique descriptive nous semblent aussi s'inscrire dans ce type de démarche. Lorsque Pierre Bourdieu et Alain Darbel travaillaient en relation avec le service d'études et de recherches du ministère de la Culture dans les années 1960, la préoccupation du ministère Malraux était de comprendre les causes de ce qui était alors perçu comme l'échec de la politique de démocratisation culturelle initiée par le ministère. L'enquête menée par Bourdieu et Darbel sur les musées européens avait pour ambition d'offrir des éléments d'explication. Même si l'on peut discuter de la nature véritablement évaluative de cette enquête, il est intéressant d'en retracer les opérations, qui sont courantes dans un certain nombre d'évaluations. La première étape est celle d'une enquête exploratoire basée sur des entretiens libres, destinée à définir les catégories pertinentes d'un ques-

tionnaire adressé à des usagers issus de divers milieux sociaux. La deuxième est celle d'un échantillonnage basé sur la fréquentation des musées étudiés. L'absence de statistiques mensuelles des visites dans certains musées et l'hétérogénéité de la compatibilité ont contraint à un travail de contrôle fastidieux : établissement de moyennes journalières, saisonnières, annuelles, puis sélection de musées représentatifs des différents types de musées, puis constitution d'un panel représentatif d'usagers et choix aléatoire des personnes interrogées. Le texte du livre *L'amour de l'art*, dont est issu l'enquête, retrace l'ensemble des phases de traitement statistique des questionnaires : premier programme d'analyse factorielle sur 1 000 questionnaires, enquêtes de vérification des dires des enquêtés sur leur temps de visite effective du musée, établissements de coefficients de correction sur cette base, puis enquête sur l'ensemble de l'échantillon, traitement par corrélations, etc. Ce qui est intéressant dans ce type d'usage de la statistique, c'est que la volonté de comprendre des pratiques suppose un travail de vérification par l'observation ou l'entretien. D'autre part, elle est mobilisée pour une description fine des pratiques, mais pas pour le travail de construction des mondes possibles à partir de cet état du monde présent. Une fois le constat posé d'une différence de pratiques selon les milieux sociaux, l'élaboration des possibles s'appuie sur une critique sociale de la monopolisation des instruments d'appropriation des biens culturels, pas sur la recherche de l'effet pur d'une variable. Comme l'a bien résumé Alain Desrosières (2003b) : « L'acteur principal de ce théâtre statistique reste le groupe social ou l'individu, ce qui distingue complètement ce théâtre de celui des méthodes économétriques ultérieures (régression logistique) qui envahiront ensuite la sociologie quantitative, et dans lesquelles les "protagonistes" (et sujets des verbes dans les phrases) sont des "variables", c'est-à-dire des entités abstraites construites sur le modèle ahistorique des sciences de la nature. C'est de ce point de vue que la sociologie empirique de Bourdieu diffère radicalement de celles d'autres courants pour lesquels la question centrale et quasiment unique est celle de "l'effet d'une variable sur une autre", question qu'il ne pose pas dans ces termes. (...) Ceci n'implique pas que Bourdieu s'interdise de parler en termes d'"effet d'une variable" (notamment la fameuse "CS du père"), mais il refuse de s'enfermer dans

la recherche des “effets purs d’une variable”, résultant de la distillation fractionnée opérée par les modèles économétriques de type LOGIT. »

Le résumé que propose Desrosières montre bien que les types de démarche ne se résument pas à un débat entre méthodes qualitatives et quantitatives mais bien à des styles de raisonnement qui peuvent cliver l’usage même d’un type de méthode.

4. LA DÉMARCHE PROCÉDURALE

Nous entendons par démarche procédurale toute démarche qui part du postulat de l’irréductibilité des valeurs pour définir des procédures de médiation ou de délibération susceptibles de faire émerger un accord entre parties intéressées à un enjeu commun. Comme nous nous centrons sur l’évaluation de l’action publique, nous excluons de notre réflexion toutes les questions désormais classiques relatives à la procéduralisation de la justice, pour nous centrer sur les procédures de débat public. Sous l’effet de la forte diffusion des travaux de Jürgen Habermas sur la démocratie délibérative, on ne compte plus les promoteurs ou les observateurs du tournant délibératif dans la conduite de l’action publique¹³. Ce qui nous intéresse plus particulièrement ici est l’usage proprement évaluatif qui peut être fait des procédures de débat comme les conférences de citoyens ou les procédures de débats publics les plus récemment organisés en France par la Commission nationale du débat public (CNDP). Nous n’ignorons pas que les procédures de débat public sont bien plus que de simples exercices d’expertises¹⁴. Mais ce qui nous intéresse ici, c’est la tendance de plus en plus affirmée à l’organisation de grands débats

13 Pour faire le point sur ces auteurs et leurs débats, nous recommandons la lecture du numéro spécial « *Debate* » de la *Revue suisse de science politique*, paru en décembre 2004.

14 Pour une vue d’ensemble de la place et du rôle des débats dans le cours de la décision publique, on se reportera avec profit aux références suivantes : Lafaye, Thévenot, 1993 ; Pourchier, 1994 ; Charlier, 1999 ; Leborgne, 1999 ; Lolive, 1999 ; Blatrix, 2000 ; Cefaï, Trom, 2001 ; Rui, 2001 ; Fourniau, 2001, 2004.

régionaux ou nationaux destinés à évaluer *ex ante*, c'est-à-dire en amont de la décision, à un moment où l'opportunité des projets peut être discutée. Les récents débats organisés par la CNDP à propos de la ligne grande vitesse PACA ou de la gestion des déchets radioactifs nous semblent aller dans ce sens. Nous faisons donc l'hypothèse qu'émerge en France un type nouveau d'évaluation *ex ante*, portée par les acteurs eux-mêmes. Il s'agit alors de comprendre le cheminement par lequel est traité l'information dans ce type de démarche. Contrairement aux démarches précédentes, l'objet n'est pas seulement de sélectionner ou de recueillir des informations, mais de les mettre en commun et de faire émerger des arguments nouveaux de la délibération. Comme le résumait bien un président de commission de débat public : « Ce que l'on cherche, ce n'est pas un sondage, ce sont des arguments pour et des arguments contre qui permettent de définir une solution qui convienne à tous¹⁵. »

La procédure de débat public institutionnalisé permet le libre accès de tous, chaque citoyen qui s'estime concerné peut participer au débat et contribuer à élargir l'éventail des mondes possibles (Fourniau, 2004). La nature des interventions au débat ne peut être anticipée et le caractère prévisible de certaines interventions ne doit pas occulter le fait que surgissent des arguments qui n'avaient pas été anticipés et qui sont pourtant techniquement recevables. Ce que soulignent les études sur la démocratie technique, c'est justement la capacité des profanes à formuler des avis scientifiquement recevables et pourtant non anticipés par les scientifiques (Callon, Barthe, Lascoumes 2003). Dans notre vocabulaire, la prise en compte est ici maximale et toutes les garanties sont données pour qu'il en soit ainsi. L'opération de prise en compte est rendue transparente et, on serait tenté de dire, exemplaire : chaque intervention est transcrite et archivée. Les choses se compliquent singulièrement lorsqu'il s'agit d'effectuer une synthèse de ces contributions hétérogènes pour en dégager les faits marquants. Les opérations de hiérarchisation ne pourront être explicitées, quels que soient les efforts

15 Philippe Marzolf, président de la commission particulière du débat public sur la ligne ferroviaire à grande vitesse PACA, interview TV à France 3 Provence, 9/05/2005.

explicitées, quels que soient les efforts déployés en ce sens par la personne en charge de la rédaction finale.

La démarche est donc totalement opposée à la démarche utilitariste, où c'est l'exact opposé qui se produit : les informations prises en compte sont limitées et le mode de hiérarchisation est protocolarisé dans les moindres détails. Baliser le chemin qui mène des débats jusqu'à des recommandations pratiques est tout simplement impossible. Le récent débat sur l'école organisé par le gouvernement Raffarin en fournit un exemple frappant : même si un luxe de précaution est pris pour que la synthèse soit exemplaire (recours à des universitaires spécialisés dans le traitement informatique des données textuelles, présence d'un conseil scientifique prestigieux, mise en ligne d'un maximum d'information relative au traitement de l'information, etc), il reste délicat d'explicitier comment on passe de 15 000 débats et de centaines d'interventions écrites – et des mètres cubes de papier associés ! – à un rapport de quelques centaines de pages. En d'autres termes, l'opération qui publiquise cette construction relève d'un travail de synthèse dont il est difficile d'explicitier les règles propres, la manière dont la synthèse de ces débats a été effectuée par Claude Thélot ne peut être explicitée. Cette limite méthodologique justifie le plus souvent la prudence des conclusions des experts. Les mondes possibles sont moins strictement hiérarchisés que présentés sous forme d'options plus ou moins souhaitables au regard des opinions exprimées au cours des consultations.

Le rapport de M. Marzolf, président de la commission particulière du débat public sur la ligne à grande vitesse PACA, est tout à fait éclairant. Le début du rapport insiste sur la prise en compte, avec l'ensemble des moyens employés pour recueillir le maximum d'avis argumentés (voir encadré).

Modalités de consultation du débat public sur la LGV PACA

(source : rapport Marzolf)

- exposition itinérante sur le débat public : 128 lieux d'exposition, 18 semaines d'exposition, 79 lieux de tractage et 39000 journaux et autant de synthèses du dossier du maître d'ouvrage tractés en marge de l'exposition (sur un total de 80 000).
- un site Internet (189 connexions quotidiennes en moyenne, 35000 visites en 6 mois) ;
- une carte T pour s'abonner à l'actualité du débat : fichier de 9620 noms collectés. Une grande partie des 7500 dossiers du maître d'ouvrage diffusés l'ont été suite à une demande par carte T ;
- un système de questions/réponses : 1200 questions posées (dont 277 avis) ;
- un journal du débat : 5 numéros, dont le 1er diffusé à 1 500 000 exemplaires, a présenté le calendrier et les modalités. Le journal n° 5 diffusé en septembre a porté à la connaissance du public la synthèse du compte rendu de la CPDP ;
- 39 réunions publiques réparties comme suit :
 - 3 réunions de lancement, à Nice, Toulon et Marseille ;
 - 3 auditions publiques pour que les acteurs fassent connaître leur point de vue au début du débat ;
 - 8 réunions thématiques sur les grands thèmes que la commission avait identifiés au début du débat : l'arc méditerranéen, les questions foncières et d'urbanisme, les enjeux d'environnement (2 réunions), le développement territorial, les transports régionaux (2 réunions) et le transport des marchandises ;
 - 5 ateliers pédagogiques : sur le financement, les solutions alternatives proposées par les acteurs et la présentation du dire d'expert commandé par la Commission nationale du débat public (3 réunions) ;
 - 17 réunions de proximité réparties sur le territoire concerné ;
 - 3 réunions de synthèse à Nice, Toulon et Marseille.

Près de 8 000 personnes ont participé à ces réunions. D'autre part, plus de 300 contributions, délibérations et motions ont été adressées à la CPDP (et mises en ligne), ainsi que 14 pétitions ayant recueilli au total plus de 26 000 signatures, présentant des avis diversifiés sur le projet. Certaines contributions ont fait l'objet de cahiers d'acteurs : en effet, 55 d'entre elles ont été sélectionnées pour leur pertinence et leur apport significatif au débat. Ces cahiers d'acteurs ont été édités en moyenne à 15 000 exemplaires chacun et c'est près de 750 000 exemplaires au total qui ont été distribués dans les réunions et adressés aux abonnés à l'actualité du débat.

On voit bien que ce type de démarche peut permettre de faire émerger les formes d'un accord sur l'état du monde présent, et sur les dilemmes dont il est porteur. Mais pour étayer un travail de construction des mondes possibles plus ordonné, il faut recourir à des opérations cognitives susceptibles de définir les critères de fermeture des options : analyse des coûts, sondages d'opinion, etc. C'est bien ce qu'illustre bien la conclusion du rapport Marzolf sur le débat relatif à la LGV PACA : « À la question, posée au débat, de l'opportunité d'une ligne nouvelle à grande vitesse dans la région PACA, la réponse est complexe : une ligne nouvelle, oui mais pas nécessairement de bout en bout ni forcément à grande vitesse (bien que cette dernière soit ressentie comme une très forte nécessité dans les Alpes-Maritimes). Et, surtout, à condition que cette ligne nouvelle soit assortie d'un vrai programme de transports régionaux, qu'elle respecte l'environnement, que les gares soient au cœur des agglomérations, qu'elle s'intègre à un vrai plan d'aménagement du territoire, et qu'elle apporte un réel progrès au quotidien (report modal, gain en temps de parcours...) ! Par conséquent, avant de répondre à la question de l'opportunité d'une LGV, les participants au débat souhaitent que le maître d'ouvrage et les élus prennent en compte leurs priorités et répondent à leurs inquiétudes. »

CONCLUSION

Si l'on suit la réflexion de Goodman sur les manières de faire des mondes, on voit bien que ce qui est en jeu dans la discussion est bien la correction (*rightness*) des versions du monde proposées par l'évaluateur. Les tenants de la démarche utilitariste insistent sur la stabilité des catégories et les propriétés logiques de leur méthode de construction des mondes possibles. Les tenants des démarches praxéologique ou procédurale insistent au contraire sur le fait que la compréhension fine des événements à partir d'une recherche approfondie d'indices et de cas exemplaires permet la reconception de l'état du monde présent. Ils jugent donc beaucoup trop sommaire la version qu'en proposent les économètres. Mais leur construction des mondes possibles ne tient pas sa force probante d'un raisonnement formellement logique. Elle la tient du fait que la version de l'état du monde présent trouve preneur au re-

gard de l'expérience commune. Les critères de jugement de la correction des versions n'étant pas les mêmes, et n'intervenant pas sur les mêmes versions mobilisées dans le raisonnement évaluatif (monde présent, mondes possibles), la controverse entre les tenants de ces démarches est sans fin. Si nous admettons avec Jean-Claude Passeron qu'« un raisonnement sociologique ne peut être de part en part ni de bout en bout un raisonnement expérimental » (Passeron, 1991), ces controverses ont peu de chance de s'apaiser grâce à l'improbable apparition d'une méthode qui mette tout le monde d'accord.

Notre apport est de qualifier plus avant l'incompatibilité entre différents styles d'évaluation en pointant l'impossibilité de combiner deux opérations, la hiérarchisation et la prise en compte, de manière transparente. Il reste à préciser si cette impossibilité est une impossibilité pratique ou logique.

Si nous essayons de faire travailler le tableau proposé, nous comprenons que ces démarches ne sont pas totalement exclusives. Il est possible de parcourir le continuum séquentiellement : compréhension des pratiques, analyse des préférences, débat, instrumentation, et vice versa (avec itérations successives éventuellement). La tendance est aujourd'hui d'expérimenter des démarches mixtes.

- La combinaison des démarches de type utilitaire et instrumental nous semble caractériser l'ensemble des opérations regroupées aujourd'hui sous le vocable générique de « nouveau management public » (Pollitt & Bouckaert, 2000). Ces opérations ont des effets politiques et sociaux importants dans la mesure où les hiérarchies produites peuvent justifier des transferts d'allocation des ressources parfois assez radicaux¹⁶.
- À l'opposé, la combinaison d'enquêtes sur les pratiques et de débats publics provoque des changements beaucoup plus incrémentaux, et tend à reproduire ce que nous appellerions volontiers une « sphère de gestion concertée ».
- Une démarche de plus en plus fréquente tend à concilier les démarches situées au centre du tableau par le recours au benchmarking, qui consiste à produire des in-

¹⁶ L'exemple le plus paradigmatique étant sans doute le train des réformes thatchériennes, bien résumé par Patrick Le Galès (2004).

dicateurs et à mettre en œuvre des changements sur la base d'études de cas visant à l'identification des meilleures pratiques (*good practices*). L'Union européenne fait un usage étendu de ce type de démarche.

- D'autres innovations plus radicales visent à concilier les démarches utilitaristes et procédurales. Le meilleur exemple est celui des sondages délibératifs mis en œuvre par James Fishkin et ses collaborateurs (Fishkin, Luskin, 1996 ; Fishkin, Luskin Jowell, 2003), qui consistent à alterner des phases de débats en conférences de citoyens et des sondages relativement élaborés. Cette démarche combinatoire permet en quelque sorte de tester les arguments décisifs et de réduire progressivement le champ des possibles.

Le choix d'une démarche est souvent suspecté de dissimuler des intérêts particuliers et d'inavouables idéologies. Les développements présentés ici n'épousent pas cette thèse souvent trop réductrice. Ils placent plus volontiers les antagonismes maintes fois constatés sur un plan logique ou pratique. Cependant, il est possible de montrer que privilégier un mode d'expertise revient à attribuer en creux une place spécifique au politique. On s'aperçoit que le rôle des autorités politiques dans la définition de la procédure d'évaluation est marginalisé aux deux pôles : celui où les mondes possibles sont fortement hiérarchisés selon un protocole scientifique, et celui où ces mondes possibles sont simplement élaborés par une opération de synthèse effectuée par une personnalité indépendante. On peut comprendre que les politiques qui souhaitent avoir prise sur le travail d'évaluation préfèrent les démarches proches des types instrumental ou praxéologique. À l'inverse, le pouvoir des évaluateurs, ou des savants pour reprendre l'appellation de Max Weber, est d'autant plus fort que l'on situe aux extrêmes du continuum. Le politique doit-il déléguer aux savants la charge de hiérarchiser les mondes possibles, quitte à fournir les critères de hiérarchisation ? Ou vaut-il mieux que l'expert se spécialise dans une fonction de prise en compte, ce qui le qualifie plutôt pour un rôle d'animateur du débat social ou de lanceur d'alerte ?

BIBLIOGRAPHIE

Michel AMIOT, *Les sociologues contre l'État*, Paris, EHESS, 1986.

Madeleine AKRICH, Michel CALLON, Bruno LATOUR, « À quoi tient le succès des innovations ? », *Annales des Mines*, 4 (29), 1988.

Lusin BAGLA-GÖKALP, « Le chercheur et son instrument : changement des systèmes de mesure et des pratiques scientifiques en mécanique des fluides », *Revue française de sociologie*, XXXVII (4), octobre-décembre 1996, pp. 537-566.

Yannick BARTHE, Michel CALLON, Pierre LASCOUMES, *Agir dans un monde incertain. Essai sur la démocratie technique*, Paris, Seuil, Coll. La couleur des idées, 2001.

Cécile BLATRIX, *La « démocratie participative », de Mai 1968 aux mobilisations anti-TGV. Processus de consolidation d'institutions sociales émergentes*, thèse de science politique, Université Paris I, janvier 2000.

Pierre BOURDIEU, Alain DARBEL, *L'amour de l'art. Les musées et leur public*, Paris, Minuit, 1966.

Valérie BOUSSARD, Delphine MERCIER, Pierre TRIPIER, *L'Aveuglement organisationnel*, Paris, CNRS, 2004.

Pierre CAHUC, André ZYLBERBERG, *Le chômage, fatalité ou nécessité ?*, Paris, Flammarion, 2004.

Daniel CEFĂI, Danny TROM (dir.), *Les formes de l'action collective. Mobilisations dans des arènes publiques*, Paris, EHESS, Raisons pratiques, 2001

Michel de CERTEAU, *L'invention du quotidien. 1. Arts de faire*, Folio essais, 1990.

Philippe CHARRETTON, Bernard FISCHESSE, « L'analyse paysagère dans la gestion des territoires », *Ingénieries - EAT*, n° 1, 1995, pp. 31-40.

Bruno CHARLIER, *La défense de l'environnement : entre espace et territoire. Géographie des conflits environnementaux déclenchés en France depuis 1974*, thèse de géographie et d'aménagement, Université de Pau et des Pays de l'Adour, 1999.

Yves CROZET, « Calcul économique et démocratie : des certitudes technocratiques au tâtonnement politique », *Cahiers d'économie politique*, n° 47, 2004, pp. 155-172.

Alain DESROSIÈRES, *La politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte, 1993.

Alain DESROSIÈRES, « Historiciser l'action publique : l'État, le marché et les statistiques », dans P. LABORIER, D. TROM, *Historicités de l'action publique*, Paris, PUF, 2003, pp. 207-221.

Alain DESROSIÈRES, « Bourdieu et les statisticiens : une rencontre improbable et ses deux héritages », dans B. ENCREVÉ, R.-M., LAGRAVE, *Travailler avec Bourdieu*, Paris, Flammarion, 2003b, pp. 209-218.

Patrice DURAN, Éric MONNIER, « Le développement de l'évaluation en France. Nécessités techniques et exigences politiques », *Revue française de science politique*, n° 2, 1992, pp. 235-262.

Patrice DURAN, « L'évaluation, au risque de son institutionnalisation », *Politiques et management public*, 20 (4), décembre 2002, introduction (pp. I-V).

James FISHKIN, Robert LUSKIN, « The deliberative poll. A reply to our critics », *The Public Perspective*, 1996, pp. 45-49.

James FISHKIN, Robert LUSKIN, Roger JOWELL, « Considered opinions : deliberative polling in Britain », *British Journal of Political Science*, n° 32, 2003, pp. 455-487.

Emmanuel FLACHAIRE, Guillaume HOLLARD, « Une approche comportementale de l'évaluation contingente », *Revue économique*, à paraître, 2006.

Joseph FONTAINE, « Évaluation et sciences sociales utiles. Raison des décideurs, offre d'expertise et usages sociaux dans quelques pratiques régionales », *Politix*, n° 36, pp. 51-71.

Joseph FONTAINE, Philippe WARIN, « Retour d'évaluations : la régionalisation des politiques publiques entre affichage et incertitude », *Pôle sud*, n° 12, mai 2000, pp. 95-112.

Jean-Michel FOURNIAU, « Mécontentement et délibération dans les conflits d'aménagement : l'expérience du débat public institutionnalisé », dans J.-P. BLAIS, C. GILLIO, J. ION (dir.), *Dynamiques associatives et cadre de vie*, Paris, Meltt-PUCA, 2001, pp. 261-301.

François FOURQUET, *Les comptes de la puissance . Une histoire de la comptabilité nationale et du plan*, Encres, recherches, 1980.

Bastien FRANÇOIS, Erik NEVEU, *Espaces publics mosaïques. Acteurs, arènes et rhétoriques des débats publics contemporains*, Rennes, Presses de l'Université de Rennes, 1999.

Ghislain GENIAUX, Guillaume HOLLARD, Claude NAPOLÉONE : « Peut-on quantifier le territoire ? » dans A. DONZEL (dir.), *Métropolisation, gouvernance et citoyenneté dans la région urbaine marseillaise*, Paris, Maisonneuve et Larose, 2001.

Olivier GODARD, « Autour des conflits à dimension environnementale – Évaluation économique et coordination dans un monde complexe », *Cahiers d'économie politique*, n° 47, 2004, pp. 127-153.

Jürgen HABERMAS, *Droit et démocratie. Entre faits et normes*, trad. franç. R. Rochlitz et C. Bouchindhomme, Paris, Gallimard, NRF essais, 1997.

Steve JACOB, Frédéric VARONE, *Évaluer l'action publique. État des lieux et perspectives en Belgique*, Gent, Academia Press, 2003.

Steve JACOB, Frédéric VARONE, « Cheminement institutionnel de l'évaluation des politiques publiques en France, en Suisse et aux Pays-Bas (1970-2003) », *Politiques et management public*, 22 (2), pp. 135-152.

Steve JACOB, Frédéric VARONE, « Institutionnalisation de l'évaluation et nouvelle gestion publique : un état des lieux comparatif », *Revue internationale de politique comparée*, 11 (2), pp. 271-292.

Marie-Christine KESSLER, Pierre LASCOUMES, Michel SETBON, Jean-Claude THOENIG, (dir.), *Évaluation des politiques publiques*, Paris, L'Harmattan, 1998.

Claudette LAFAYE, Laurent THÉVENOT, « Une justification écologique ? Conflits dans l'aménagement de la nature », *Revue française de sociologie*, 34 (4), 1993, pp. 495-524.

Pierre LASCOUMES, « Pratiques et modèles de l'évaluation », dans M.-C. KESSLER, P. LASCOUMES, M. SETBON, J.-C. THOENIG, *Évaluation des politiques publiques*, Paris, L'Harmattan, Coll. Logiques politiques, 1998.

Pierre LASCOUMES, Michel SETBON, *L'évaluation pluraliste des politiques publiques : enjeux, pratiques, produits*, GAPP-CNRS-Commissariat Général au Plan, 1995.

Pierre LASCOUMES, Patrick LE GALÈS, *Gouverner par les instruments*, Paris, Presses de Sciences Po, Coll. Gouvernances, 2004.

Bruno LATOUR, *Politiques de la nature. Comment faire entrer la science en démocratie*, Paris, La Découverte, Armillaire, 1999.

Mathieu LEBORGNE, *De quelques caractères associatifs. La mobilisation associative entre l' élu local et le citoyen lors du débat public relatif au projet de ligne THT entre Boute (Var) et Carros (Alpes-Maritimes)*, Marseille, ministère de l'Environnement et de l'Aménagement du territoire et SHADYC (EHESS), janvier 1999.

Olivier LENAY, « De la concrétisation d'un outil de gestion : le rôle du PMSI dans la régulation du système hospitalier », *Politiques et management public*, 18 (3), 2000.

Jacques LOLIVE, *Les contestations du TGV Méditerranée*, Paris, L'Harmattan, 1999.

Yves MÉNY, Jean-Claude THOENIG, *Politiques publiques*, Paris, PUF, 1989.

Jean-Claude MOISDON, « Quelle est la valeur de ton point ISA ? Nouveaux outils de gestion et de régulation dans le système hospitalier français », *Sociologie du travail*, 42 (1), 2000, pp. 31-49.

Jean-Pierre NICOLAS, Fabien DUPREZ, Sandrine DURAND, Fabrice POISSON, Pierre-Louis AUBERT, Mireille CHIRON, Yves CROZET, Jacques LAMBERT, *10 ans de monétarisation des effets locaux de la pollution atmosphérique*, Synthèse Inrets n° 46, 2002.

Frédérique PALLEZ, « De la mesure dans un service public régalién. Peut-on et faut-il quantifier la charge de travail des magistrats ? », *Politiques et management public*, 18 (4), décembre 2000, pp. 91-118.

Jean-Claude PASSERON, *Le raisonnement sociologique*, Paris, Nathan, 1991.

Jean-Claude PASSERON, « Statique et dynamique. Caractériser, expliquer, comprendre le changement », *Revue européenne des sciences sociales*, XXXVI (110), 1998, pp. 199-263.

Jean-Claude PASSERON, Louis-André GÉRARD-VARET (dir.), *Le modèle et l'enquête. Les usages du principe de rationalité en sciences sociales*, Paris, EHESS, 1995.

Thomas PIKETTY, « L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises : une estimation à partir du panel primaire 1997 », rapport en ligne du pôle Jourdan, EHESS, 2004, 71 p. <http://www.jourdan.ens.fr/piketty/fichiers/public/Piketty2004b.pdf>

Christopher POLLITT, Geert BOUCKAERT, *Public Management Reform. A Comparative Analysis*, Oxford, Oxford University Press, 2000.

Éric POURCHIER, « De la CARDE à la FARE-Sud : le mouvement de contestation du TGV Méditerranée », dans Alain HAYOT (dir.), *Ville et citoyenneté*, Villes et territoires méditerranéens, Montpellier, 1994.

Sandrine RUI, *La démocratie en débat. Les citoyens face à l'action publique*, Paris, Armand Colin, 2004.

Sandrine RUI, *L'expérience démocratique. Observation et analyse de l'implication des citoyens dans les procédures de concertation relatives à la mise en œuvre de projets d'infrastructure et d'aménagement*, thèse de sociologie, Université de Bordeaux II, janvier 2001.

Vincent SPENLEHAUER, *L'évaluation de politiques, usages sociaux. Trois études de cas d'évaluation*, Paris, L'Harmattan, 1995.

Vincent SPENLEHAUER, « Quel développement, quelle professionnalisation de l'évaluation ? », *Pouvoirs locaux*, n° 38, 1998.

Vincent SPENLEHAUER, « Intelligence gouvernementale et sciences sociales », *Politix*, n° 48, décembre 1999, pp. 95-128.

Vincent SPENLEHAUER, « L'évaluation au service des conseils régionaux », *Sociologie du travail*, n° 42, 2000, pp. 245-263.

Jean-Claude THOENIG, « L'évaluation en actes. Leçons et perspectives », *Politiques et management public*, 20 (4), 2002, pp. 33-50.

Bruno VAN DER LINDEN, « L'effet des formations professionnelles de chômeurs : de l'impact sur les individus à l'impact macroéconomique », *Bulletin de l'IRES*, Bruxelles, n° 227, 2001 (<http://www.ires.ucl.ac.be>)

Antoine VION, « Construire une typologie de *policy takers* : une opération logique et heuristique », *Revue internationale de politique comparée*, 4^e trimestre, décembre 2004.