

Performance, incitations et gestion publique

Rapport

*Dominique Bureau
et Michel Mougeot*

Commentaire

Philippe Mongin

Compléments

*Frédéric Bousquet, Alain Coulomb, Jean-Luc Gaffard,
Sylvie Hel-Thelier et Martine Perbet*

*Réalisé en PAO au Conseil d'Analyse Économique
par Christine Carl*

© La Documentation française. Paris, 2007 - ISBN : 978-2-11-006636-7

« En application de la loi du 11 mars 1957 (article 41) et du Code de la propriété intellectuelle du 1er juillet 1992, toute reproduction partielle ou totale à usage collectif de la présente publication est strictement interdite sans l'autorisation expresse de l'éditeur.

Il est rappelé à cet égard que l'usage abusif de la photocopie met en danger l'équilibre économique des circuits du livre. »

Sommaire

Introduction	5
<i>Christian de Boissieu</i>	

RAPPORT

Mesure de la performance et incitations dans la gestion publique	7
<i>Dominique Bureau et Michel Mougeot</i>	

<i>Introduction</i>	7
---------------------------	---

<i>1. Démarche de performance et indicateurs</i>	10
1.1. Le contexte	10
1.2. La mesure de la performance : problèmes de méthodes	18
<i>2. Incitations et gestion publique</i>	23
2.1. La vision économique de l'intervention publique aujourd'hui	23
2.2. Contraintes informationnelles, rentes et incitations	30
<i>3. L'organisation de l'évaluation</i>	37
3.1. Éléments d'analyse	38
3.2. Enseignements et propositions	43
<i>4. Évaluation des performances et réforme de l'État</i>	49
4.1. La possibilité de gérer : organisation des agences et discrétion laissée aux agents	49
4.2. Une perspective pluriannuelle contractualisée	52
4.3. Un marché interne du travail plus flexible	53
<i>5. Conclusion : des services et agences rendant compte des services rendus dans des conditions permettant la comparaison des performances</i>	54

COMMENTAIRE

<i>Philippe Mongin</i>	61
------------------------------	----

COMPLÉMENTS

A. Mesure de la performance et gestion hospitalière	71
<i>Frédéric Bousquet et Alain Coulomb</i>	
B. L'évaluation de l'enseignement supérieur et de la recherche : le cas d'école de la réforme « LMD »	79
<i>Jean-Luc Gaffard</i>	
C. La mesure de la production et de la productivité des administrations publiques : la mission Atkinson au Royaume-Uni	85
<i>Sylvie Hel-Thelie</i>	
D. L'évaluation en France : retour d'expérience	101
<i>Martine Perbet</i>	
E. La décision publique au Royaume-Uni	113
<i>Martine Perbet</i>	
RÉSUMÉ	127
SUMMARY	131

Introduction

La réforme budgétaire en France est incarnée par la LOLF (loi organique sur les lois de finances), qui implique de se préoccuper désormais de l'efficacité de la dépense publique. À ce titre, il convient donc de se donner les moyens de mesurer cette efficacité et d'en tirer toutes les conséquences pour le fonctionnement des administrations, l'organisation de l'État, les règles de la fonction publique...

Le rapport qui suit met en lumière et relie d'une part les méthodes d'évaluation des performances du secteur public, d'autre part les systèmes d'incitation à destination des agents publics, encore peu développés et qu'il va falloir promouvoir ne serait-ce que pour mieux articuler les considérations d'efficience et d'équité.

L'élaboration d'une batterie d'indicateurs de performance adaptés et fiables est une condition première du succès, pas encore garanti, de la LOLF. Le rapport insiste sur les biais introduits par des indicateurs partiels, sur l'importance de l'évaluation de la qualité des services collectifs et sur l'intérêt des comparaisons internationales (*benchmarking*). L'indépendance et la crédibilité des évaluateurs sont fondamentales, exigeant l'absence de conflits d'intérêts caractérisés et suscitant une méfiance à l'égard d'instances centralisées bénéficiant d'un monopole.

La France doit clairement développer sa culture et sa pratique d'expérimentation et d'évaluation pour les politiques publiques. Pour ce qui touche à la production des services collectifs, le système des agences, pratiqué chez certains de nos voisins (en particulier le Royaume-Uni), constitue dans certains cas la bonne réponse, et une application directe des modèles de délégation (principal/agent). Quel que soit le degré effectif de déconcentration ou de décentralisation, il faut mettre en place un système d'incitation et d'intéressement des personnels qui aille dans le sens d'une efficacité accrue du secteur public, qu'il s'agisse de l'État, des collectivités locales ou de la Sécurité sociale.

L'application de ces principes à deux domaines phares, les hôpitaux et les universités, permet de donner de la chair aux propositions concrètes pour améliorer la gestion publique.

Ce rapport a bénéficié du soutien efficace de Fabrice Lenseigne, conseiller scientifique au CAE. Il a fait l'objet d'une présentation à Jean-François Copé, ministre délégué au Budget et à la Réforme de l'État, le 18 janvier 2007.

Christian de Boissieu
Président délégué du Conseil d'analyse économique

Mesure de la performance et incitations dans la gestion publique

Dominique Bureau

*Directeur des Affaires économiques et internationales,
Ministère des Transports, de l'Équipement, du Tourisme et de la Mer*

Michel Mougeot

Professeur à l'Université de Franche-Comté

Introduction

La mesure de la performance dans la gestion publique répond à un impératif démocratique, reconnu par les articles 14 et 15 de la Déclaration des Droits de l'homme et du citoyen, qui établissent, d'une part, que « tous les citoyens ont le droit de constater par eux-mêmes ou par leurs représentants, la nécessité de la contribution publique (...) et d'en suivre l'emploi (...) » et, d'autre part, que « la société a le droit de demander compte à tout agent public de son administration ».

La question de la performance dans la gestion publique est particulièrement aiguë aujourd'hui. La fin du vingtième siècle a été marquée en effet par l'échec des économies planifiées, et par la remise en cause, dans les économies développées, des modalités de l'intervention publique, notamment de régulation des industries de réseau. Ces deux processus ont en commun d'avoir ébranlé les idées en matière d'intervention publique. La possibilité d'une coordination administrée globale se trouve désormais écar-

(*) Les auteurs remercient, sans les impliquer – car la prise en compte de leurs observations aurait mérité en général un approfondissement beaucoup plus consistant –, Paul Champsaur, P. Delage, H. Eyssartier et Jean Tirole pour leurs remarques sur les versions préliminaires de ce rapport. Cette version tient compte aussi du débat au CAE le 6 novembre 2006, notamment des remarques de Jean-Michel Charpin, Jacques Delpla, Jean-Paul Fitoussi et Philippe Trainar, et bien sûr du rapporteur Philippe Mongin.

tée. La réflexion sur la régulation des monopoles naturels a, par ailleurs, introduit un nouveau bilan à effectuer, entre les imperfections du marché, que l'on veut corriger, et les imperfections des mécanismes de coordination administrés, que l'on ne peut négliger.

Ainsi, l'approche économique de la gestion publique comporte maintenant deux dimensions. La première, traditionnelle, de l'évaluation des politiques, consiste à s'assurer que les services fournis ou les réglementations procurent des bénéfices en rapport avec leur coût. La seconde porte sur le choix des modalités de l'intervention publique et, si c'est une offre publique qui est retenue, sur les conditions d'organisation et d'efficacité dans lesquelles celle-ci est fournie.

À cet égard, l'expérience accumulée par les régulateurs sectoriels, comme tutelle de monopoles publics, ou plus récemment, comme régulateurs de l'accès aux infrastructures essentielles ou de la fourniture d'obligations de service public, souligne à quel point la disponibilité d'informations pertinentes sur la performance conditionne l'efficacité de la régulation publique.

Or, comme le soulignent Alain Lambert et Didier Migaud (2005), en France, la gestion de l'État n'a jamais été orientée vers les résultats. Ces questions émergent donc dans notre pays, compte tenu du poids de l'intervention publique dans l'économie tel que le reflète le niveau de nos prélèvements obligatoires. Le besoin de moderniser l'État, en y développant la gestion par la performance, se trouve au cœur de la réforme budgétaire engagée avec la loi de finances du 1^{er} août 2001 (LOLF). Comme il n'y a pas de performances sans indicateurs appropriés, la définition de ceux-ci a retenu l'attention. Ceci a notamment conduit à la publication d'un guide méthodologique élaboré en commun par le ministère de l'Économie et de l'Industrie (Direction de la réforme budgétaire), avec les Commissions des finances de l'Assemblée nationale et du Sénat, la Cour des comptes et le Comité interministériel d'audit des programmes. Ce guide (2004) recense les critères à satisfaire pour construire de « bons » indicateurs : pertinents, utiles, solides et vérifiables. En pratique, beaucoup de progrès doivent cependant être réalisés encore pour que les indicateurs disponibles satisfassent pleinement ces critères, comme le constatait le rapport d'information sur les indicateurs de performance de la LOLF de la Commission des finances du Sénat (2005).

L'enjeu est d'importance car l'utilisation d'indicateurs imparfaits dans des mécanismes d'intéressement salarial ou d'allocation des moyens budgétaires risque d'entraîner des effets indésirables, qui pourraient délégitimer le processus de réforme engagé, s'ils conduisent à porter au pinacle des agents ayant su utiliser les failles de ces mécanismes, ou s'ils favorisent une dégradation de l'offre de services utiles, mais mal pris en compte par les systèmes d'indicateurs. Un autre risque est l'accumulation de travaux d'expertise coûteux n'aboutissant *in fine* qu'à confirmer l'insuffisance de résultats, mais visant à la rendre « compréhensible » dans une logique compassionnelle.

Pour surmonter ces obstacles, les mécanismes de gestion pour la performance doivent être fondés sur une compréhension profonde des problèmes d'information et d'incitations auxquels est confrontée la gestion publique et des conditions qui permettraient de les alléger. Notre objectif avec ce rapport est de fournir des points de repère économiques en ce sens, pour compléter les approches juridiques, sociologiques et managériales qui prévalent habituellement en ces matières. À cet égard, l'apport de l'économiste est, de notre point de vue, d'avoir en perspective un objectif explicite d'intérêt social, et une explicitation des contraintes informationnelles à prendre en compte. Du point de vue conceptuel c'est donc la théorie des incitations, développée depuis vingt-cinq ans, notamment sous l'impulsion de l'école toulousaine (autour de Jean-Jacques Laffont et Jean Tirole) qui constituera notre référence centrale. Notre rapport se présente ainsi comme le prolongement de celui de J.-J. Laffont (1999).

Toutefois, le parti qui a été retenu ici consiste à partir de l'expérience, et des difficultés rencontrées par les praticiens pour élaborer ces indicateurs. Ceci conduira à introduire de manière anticipée certaines idées, par rapport à la démarche la plus logique qui aurait consisté à rappeler les problèmes d'incitations dans les organisations en général, et l'impact de leur prise en compte pour la gestion publique, avant d'aborder les problèmes plus spécifiques de mesure de la performance publique.

De même, il convient de souligner que la thématique des indicateurs que nous explorons ici est plus étroite que celle de « l'évaluation », qui englobe (et ne distingue probablement pas assez) les différents niveaux : de la légitimation, en termes de bilan coûts-avantages, du niveau de l'action publique ; du choix du mode d'intervention retenu ; du partage des rôles entre public et privé, et de l'efficacité de l'organisation de l'État pour remplir ses missions. Notre analyse se focalisera principalement sur ce dernier aspect, dont les liens avec les autres éléments seront cependant précisés au cours de l'exposé.

Tout d'abord, nous rappellerons le contexte institutionnel dans lequel s'inscrit le développement de la gestion publique par la performance. Ceci nous amènera à qualifier, d'un point de vue économique, les problèmes rencontrés dans l'élaboration des systèmes d'indicateurs.

Partant de l'état de nos instruments, tel qu'on peut l'apprécier à la fois au niveau macroéconomique et à celui des secteurs les plus concernés – santé, éducation, notamment –, on montrera comment les difficultés rencontrées s'interprètent dans ce cadre d'analyse microéconomique, et donc qu'une approche trop technocratique de la réforme de l'État, ignorant cette dimension, risque de retrouver les écueils de la planification centralisée dont les objectifs apparaissent très vite irréalistes ou non contraignants, pour des raisons que nous rappellerons. On précisera ensuite comment les indicateurs doivent être construits avec une triple perspective de supervision, de concurrence par comparaison, et de diffusion de l'information pour éviter les collusions de groupes d'intérêt contre l'intérêt général. On esquissera enfin des propositions, dont l'idée générale est que plus de performance

nécessite plus de capacité de gestion, qui doit aller de pair avec plus d'obligation de « rendre compte ». On en précisera enfin les mesures d'accompagnement pour en assurer l'efficacité.

Dit autrement, nous commencerons par montrer que la gestion publique rencontre des problèmes de mesure de la performance particulièrement aigus, qui exacerbent les problèmes d'incitations à résoudre, par rapport à ce qu'ils sont dans d'autres organisations, et que ceci plaide pour des niveaux d'incitations « intermédiaires ». On rappellera plus généralement comment ces contraintes informationnelles ont amené les économistes à revoir leurs conceptions de l'intervention publique. On reviendra alors sur le fait que le processus d'évaluation est lui-même soumis à des contraintes d'incitations, et sur l'articulation entre évaluation et gestion dans la réforme de l'État.

Bien évidemment, on ne trouvera pas ici, fourni clef en mains, de système explicite d'indicateurs. En revanche, on espère convaincre de deux points fondamentaux. D'une part l'efficacité de ces indicateurs nécessite qu'ils soient conçus dans une perspective microéconomique, de mise en ligne des objectifs des agents ou agences avec l'intérêt général. D'autre part une telle ambition n'est pas inaccessible. En d'autres termes, on espère ainsi contribuer à un processus qui ferait que les recommandations en faveur de l'évaluation systématique de l'action publique ne seraient plus un vœu pieux des rapports sur la Finances Publiques.

1. Démarche de performance et indicateurs

[La présentation des actions de l'État est faite au regard] « des coûts associés, des objectifs poursuivis, des résultats obtenus et attendus pour les années à venir mesurés au moyen d'indicateurs précis dont le choix est justifié ».

Art. 51 de la loi organique relative aux lois de Finances du 1^{er} août 2001.

L'objectif poursuivi avec la nouvelle architecture budgétaire de la LOLF est de faire évoluer la gestion publique vers la performance, sur la base d'une stratégie, d'objectifs, et d'indicateurs précis pour guider et évaluer les politiques. Comme le rappelle le guide méthodologique cité ci-dessus, l'ambition affichée est que cette définition d'objectifs, et la mesure des résultats associés améliorent non seulement le processus de décision budgétaire, mais aussi la gestion même des administrations, ce dernier élément étant présenté comme allant de soi. Nous verrons cependant que la mise en place d'indicateurs globaux dans le cadre de la LOLF demeure loin d'être suffisante pour atteindre cet objectif.

1.1. Le contexte

1.1.1. La démarche de performance

Afin d'organiser le passage d'une logique de moyens à une logique de résultats, la LOLF a réorganisé la nomenclature budgétaire en missions, programmes et actions correspondant à la finalité des dépenses. Par ailleurs, le vote du budget est désormais accompagné de la définition d'objectifs et de cibles de résultats à atteindre qui ensuite doivent être commentés et analysés dans le cadre des rapports annuels de performances, annexés au projet de lois de règlement. Simultanément, les administrations doivent mettre un dispositif de pilotage par les performances. Enfin, un dispositif de contrôle de la pertinence et de la qualité des informations est établi, qui implique le Parlement, la Cour des Comptes, et le Gouvernement, au travers du comité interministériel d'audit des programmes.

Le document précité détaille les principes à suivre pour établir ces objectifs et indicateurs. Il souligne que les objectifs doivent couvrir trois dimensions :

- répondre aux attentes des citoyens, en informant sur l'impact socio-économique des politiques ;
- satisfaire les préoccupations de qualité de service de l'utilisateur ;
- fournir, pour le contribuable, des éléments sur l'efficacité de la gestion. S'agissant des *indicateurs*, il est rappelé que ceux-ci doivent :
- être cohérents avec les objectifs ; concerner une part substantielle des résultats ;
- permettre de porter un jugement.

Le développement d'une telle démarche s'inscrit dans une tendance commune, que l'on rencontre aux États-Unis et au Royaume-Uni et plus généralement dans les pays de l'OCDE.

Dans cette perspective Kravchuk et Schack (1996) avaient établi les dix principes suivants pour la conception des systèmes de mesure de la performance publique :

- formuler une mission claire et cohérente ;
- développer une stratégie de mesure explicite ;
- impliquer des utilisateurs clés dans la phase de création et de développement ;
- rationaliser la structure des programmes avant le début des mesures ;
- développer divers ensembles de mesures pour différents utilisateurs ;
- prendre en compte l'avis des clients tout au long du processus ;
- fournir à chaque utilisateur des renseignements suffisants ;
- revoir et réviser régulièrement les mesures ;
- prendre en compte toutes les difficultés ;
- éviter l'accumulation excessive d'informations.

Aujourd'hui, ce processus tend même à déborder sur la comptabilité nationale (Hel-Thelier, 2007). Au Royaume-Uni, le rapport Atkinson (2004)

considère que l'enjeu de la performance publique est suffisamment important pour que les principes de la comptabilité nationale eux-mêmes soient revus. La mesure de la gestion publique bascule ainsi progressivement, et de manière irréversible, de l'évaluation des moyens au coût des facteurs, vers celle de l'activité (au moins), l'objectif étant progressivement d'appréhender des résultats (« *outputs* » et « *outcomes* »).

Le besoin d'objectivation de la performance publique est ainsi reconnu et se concrétise par la mise en place d'un centre de mesure dédié. L'accent est mis sur l'importance des comparaisons entre l'évolution de la productivité publique et celle de l'ensemble de l'économie. Tony Atkinson insiste, par ailleurs, sur la nécessité d'ajuster les indicateurs sur la qualité, d'aller vers une couverture exhaustive des services et des disparités régionales, ainsi que de tenir compte des complémentarités entre services publics et secteur privé, le rôle de l'intervention publique étant souvent de fournir l'infrastructure matérielle (transports), immatérielle (éducation, recherche), ou institutionnelle nécessaire à l'économie.

Ces perspectives ne sauraient nous faire oublier, cependant, que les indicateurs calculés à cette fin demeurent très frustes. De plus, cette appréciation vaut non seulement pour le niveau très agrégé, mais concerne en fait l'ensemble des systèmes de mesure de la performance publique (cf. Propper et Wilson, 2003). Ces auteurs suggèrent que ces mesures de performances devraient donc être utilisées seulement en tant qu'indicateurs, et non pas en tant qu'outils de gestion. Ils rappellent que les organismes publics existent justement parce qu'il y a des buts contradictoires parmi les parties prenantes qu'ils représentent et qu'en conséquence il n'est pas raisonnable de penser qu'un seul ensemble de mesures puisse résoudre leurs problèmes de gouvernance.

Si l'on considère l'état du processus en France, il apparaît encore limité et fragile, comme l'a souligné la Commission d'évaluation des stratégies ministérielles de réforme. Le processus mobilise des ressources considérables pour des gains de performances qui demeurent encore minces. La complexité des instruments de mesure de performance qui seraient nécessaires conduit souvent à se rabattre sur des « *proxys* », dont l'usage pourrait vite se révéler contre-productif s'ils valorisent trop certains services. Par ailleurs, le souci de redéfinir simultanément les missions et l'organisation des services exacerbe parfois les conflits d'autorité (entre responsables administratifs et financiers et responsables opérationnels ; entre ministères techniques et régaliens ; entre tutelles et établissements publics, etc.), ce qui retarde la mise en place des instruments de mesure de la performance.

1.1.2. Retour d'expérience

Il ne s'agit pas ici de jouer les Cassandra mais de prendre la mesure du travail à réaliser. Ce n'est pas en effet la première fois que l'on cherche à introduire de l'efficacité dans notre gestion publique. Le développement de la rationalisation des choix budgétaires (RCB) à la fin des années soixante procédait de cette démarche, et pouvait s'appuyer déjà sur des concepts

d'économie publique bien établis. Développés après-guerre, ceux-ci n'ont d'ailleurs pas fondamentalement changé.

Certes, on peut objecter que la RCB a alors été handicapée :

- par l'absence concomitante de réforme de la procédure budgétaire pour intégrer ces méthodes ;
- et au niveau conceptuel, par une analyse économique focalisée sur l'évaluation coûts-avantages des politiques dans une perspective de planification, qui laissait de côté les problèmes d'organisation, dont l'analyse économique, elle, n'était pas encore disponible.

Cependant les espoirs mis depuis dans le développement de systèmes de mesure de la performance dans différents secteurs se sont révélés, souvent, tout aussi déçus.

L'examen de la situation des hôpitaux et de celle des universités apporte à cet égard des éclairages complémentaires.

Dans le cas des hôpitaux (*cf.* le complément de Frédéric Bousquet et d'Alain Coulomb), ce processus a buté sur la question de l'articulation entre le secteur public et le secteur privé. Celle-ci apparaît d'autant plus importante que leurs modes d'organisation et de fonctionnement génèrent des incitations très différentes, et que l'on assiste à des tendances à leur spécialisation, très fortes qui mériteraient d'être évaluées en termes « d'intérêt social ». On constate cependant « l'absence troublante, pour ne pas dire scandaleuse, d'études de mesures rigoureuses et objectives, qui a accompagné un débat stérile depuis près de deux décennies » sur la notion de performance (alors qu'il ne s'agit que de s'assurer que les services offerts sont en rapport avec les ressources utilisées et que celles-ci sont utilisées au mieux !), et sur l'efficacité comparée des deux secteurs.

On a assisté ainsi à une véritable collusion pour conserver deux échelles de mesure, construites par ailleurs avec des groupes homogènes de malades trop larges. Pourtant quelques années suffisent normalement pour définir les bases d'un tel système d'indicateurs, même si les « *case-mix* » – c'est-à-dire les pondérations permettant de corriger les indicateurs, pour tenir compte de la lourdeur relative des pathologies traitées – doivent ensuite être perfectionnés, de manière à ce que les débats sur la performance ne soient pas pollués par des problèmes méthodologiques que l'on peut résoudre. Par ailleurs, il semble que l'on ait, dans ce cadre, plus aidé les moins performants que cherché à récompenser les « bons », les points « ISA » servant de base à la rémunération des établissements étant ajustés individuellement. À noter enfin la faible diffusion (tous les quatre ans) de l'information acquise dans le processus d'accréditation des établissements.

Frédéric Bousquet et Alain Coulomb estiment pourtant que le fait même de mesurer la performance permet de l'améliorer, d'où l'urgence qu'il y a à disposer d'un instrument de mesure de référence valide et fiable. L'introduction de la qualité dans le débat leur paraît un moyen puissant de renfor-

cer l'efficacité et l'équité du système. Personne ne peut en effet récuser cette dimension, dont le contrôle est par ailleurs d'autant plus important que l'on développe des mécanismes de rémunération forfaitisés visant à maîtriser les coûts, comme cela sera développé ci-dessous. Le thème de la qualité est aussi un moyen de créer des collusions favorables avec le public.

Ces constatations trouvent un écho direct dans le monde universitaire. À cet égard Bernard Belloc, ancien vice-président de la Conférence des présidents d'Université, estime en effet, qu'en France « trop d'évaluation tue l'évaluation » (surtout lorsqu'elle n'a pas de conséquence !) et que l'on s'ingénie trop souvent à éviter les systèmes d'indicateurs, tels que les publications des chercheurs (et notamment les facteurs d'impact), par exemple, qui ont été développés et perfectionnés partout ailleurs. Il observe aussi que les incitations en découlant fonctionnent souvent à l'envers (« vous êtes bons, donc vous n'avez pas besoin d'argent ! »).

1.1.3. Les indicateurs de la LOLF

Peut-on être plus optimiste quant à l'essor des mesures de performances associées à la LOLF ? Clairement, il est trop tôt pour répondre à cette question. L'essentiel de la littérature dont on dispose sur ce sujet demeure d'ailleurs interne à l'État, et il est normal, que sur un processus en chantier, celle-ci soit biaisée, selon les cas, dans un sens critique ou au contraire « intentionnel », qui participe à la gestion du projet.

L'étude de Lacaze (2005), qui examine l'application de la LOLF au cas du management de la force publique est donc précieuse, car il s'agit là d'un point de vue externe. Celui-ci souligne tout d'abord la complexité de la tâche : « comment évaluer, en effet, la performance des services de sécurité alors que l'effet recherché est atteint lorsque certains actes délictueux sont déjoués et que, *in fine*, il ne se passe rien ? » Certes, on objectera que le cas étudié est parmi les plus difficiles. Mais il est aussi incontournable, sachant que les questions de sécurité et celles de respect du droit sont au cœur d'une part très importante de l'action publique. L'étude constate en fait que la LOLF a permis un effort de formulation des missions, et que la définition des actions – ici, celles de la Police nationale – les reflète assez fidèlement. En revanche, l'analyse est beaucoup plus critique sur les objectifs et indicateurs, qui pour l'auteur :

- constituent une réduction et une parcellisation du champ des actions, et ne rendent donc pas compte entièrement des missions effectivement réalisées ;
- préjugent des moyens adaptés à mettre en œuvre pour accomplir une action. L'optimisation de la capacité opérationnelle, par exemple, n'est qu'un moyen parmi d'autres de maîtriser les troubles de l'ordre public ;
- focalisent excessivement la mesure, l'indice de disponibilité des forces et le taux d'encadrement ne constituant qu'une approche partielle de cette optimisation, si l'on poursuit l'exemple précédent.

Lacaze pointe alors la possibilité d'effets pervers si les gestionnaires ne sont jugés que sur ces indicateurs partiels, car ils risquent, notamment, de délaisser les tâches qui nécessitent une coordination entre services. Par ailleurs, il signale que le jugement sera inéquitable si l'évolution de l'indicateur dépend trop de facteurs externes d'environnement, ou si le gestionnaire est soumis à des contraintes (y compris administratives ou statutaires) qui font qu'il n'a de prise que sur une partie de l'indicateur. De manière générale, il émet beaucoup de réserves vis-à-vis de l'accent mis sur la performance quantifiable, et souligne l'importance des questions de valeurs et de déontologie. Malgré ces critiques, il considère que ces effets pervers pourraient être limités, en élargissant notamment la gamme des indicateurs.

Le dernier point est techniquement peu contestable. Le souci de développer les systèmes d'information internes, pour disposer d'indicateurs chiffrés, conduit sans doute à être trop restrictif, et la panoplie des systèmes de mesure, pourrait être élargie dans au moins deux dimensions :

- les approches en termes de « certification » type ISO 9000, qui permettent d'assurer un contrôle qualité des procédures et par là des services ;
- le recours périodique à des enquêtes auprès du public et des usagers, étant rappelé que celles-ci doivent être conçues rigoureusement pour que les réponses reflètent bien leurs préférences.

Les autres questions soulevées sont plus fondamentales. Ce qui est en effet suggéré au fond, c'est que les systèmes mis en place relèvent parfois plus du suivi d'un plan d'action que de la mesure économique de la performance. Bien évidemment ceci n'ôte rien à l'importance de tels plans d'action, qui sont déterminants pour faire évoluer l'administration. Mais la gestion publique doit considérer à la fois l'agenda gouvernemental immédiat, et des horizons plus longs. Du point de vue économique, ces questions sont au cœur de l'analyse en termes « multitâches et multi-mandants » qui suit.

Celle-ci sera posée d'abord en termes relativement abstraits. Pour autant, les questions sous-jacentes sont bien intuitives. La performance d'un service hospitalier ne peut se limiter à la maîtrise du risque de maladies nosocomiales, même si celles-ci appellent évidemment une réaction vigoureuse. De même, la gestion du risque inondations ne se limite pas à la prescription de plans de prévention. Enfin, le développement de nouvelles matières dans l'enseignement ne peut se faire au détriment des apprentissages fondamentaux, etc.

La mesure de la performance publique ne peut ignorer ces différentes dimensions, même lorsque celles-ci ne sont pas véritablement contradictoires. En effet, toute faiblesse à cet égard – et ceci constitue la seconde ligne directrice du rapport – a des conséquences immédiates, au travers des modifications de comportement des agents qu'elles induisent (« incitations »). Certes, la multiplicité des tâches ou des mandants est souvent exagérée, ou mise en avant par ceux qui souhaitent échapper aux contraintes de performance. Par ailleurs, ces conflits d'objectifs peuvent souvent être allégés. Mais il faut justement pour cela en avoir compris la nature.

1.1.4. Performance administrative et action gouvernementale

Les observations précédentes conduisent donc à examiner le problème d'articulation qui se pose entre la mesure de la performance des services publics et des systèmes d'indicateurs limités à la réalisation de plans d'action.

Comme cela a été suggéré ci-dessus, l'écart entre les deux approches réside d'abord dans la nécessité, pour mesurer correctement la performance d'un service, de prendre en compte des objectifs plus lointains. En d'autres termes, il faut évaluer comment sont remplies les missions de long terme du service public. Du point de vue technique, ces éléments seront en général délicats à mesurer, comme le sont, dans les domaines des entreprises, la valeur des immobilisations correspondant à la recherche-développement et, *a fortiori*, les éléments de réputation qui concourent à la valeur d'une firme. Dans ce cas, cette difficulté « technique » n'est pas considérée comme un obstacle pour engager de telles évaluations, qui sont nécessaires pour les actionnaires de l'entreprise, mais pousse au contraire à rechercher les méthodes plus appropriées de valorisation des risques ou actifs concernés.

Cependant, et c'est le point crucial, ce problème d'évaluation est associé ici à un problème de gouvernance particulièrement aigu, car la poursuite de ces missions de long terme et leur mise en œuvre ne ressortissent pas seulement à la mise en œuvre de l'agenda gouvernemental « immédiat » et concernent de multiples parties prenantes. Il faut donc réfléchir au degré « d'indépendance » des services ou opérateurs publics, qui serait souhaitable pour bien les réaliser. Si ce sujet apparaît très médiatique lorsqu'il concerne quelques domaines sensibles comme l'instruction judiciaire ou l'inspection du travail, il apparaît en fait peu débattu dans sa généralité. Ainsi des principes d'autonomie des entreprises publiques avaient été posés par le rapport Nora, il y a quarante ans maintenant. Mais ils n'ont pas donné lieu ensuite à la construction d'une doctrine opérationnelle pour leur mise en œuvre, si bien que l'on assiste toujours au même tâtonnement dans l'organisation des relations entre les tutelles et leurs établissements publics. Et on assiste aux mêmes incertitudes quant au sens à donner à la constitution des régulateurs indépendants dans le domaine des grands réseaux (télécoms, électricité, audiovisuel, etc.), dont les tâches apparaissent d'ailleurs très variables d'un secteur à l'autre.

Les termes d'autonomie ou d'indépendance qui ont été employés ci-dessus sont sans doute malheureux et ont probablement bloqué la réflexion sur ces sujets, qui serait sans doute mieux posée en termes de degré de « séparation », et de modalités appropriées pour « équilibrer » les relations entre les différentes composantes de l'organisation des services publics.

Il n'en demeure pas moins que le public est très ambivalent sur ces sujets. D'un côté, il est attaché à l'idée d'un État « impartial », donc ayant nécessairement une certaine autonomie. De l'autre, l'idée économique qui sous tend l'autonomie des banques centrales par exemple (à savoir que le

manque de crédibilité des gouvernements à ne pas recourir au quotidien à une politique exagérément inflationniste, rend bénéfique de confier la politique monétaire à un agent plus « conservateur » que la société, pour réduire ce biais inflationniste) demeure perçue comme provocatrice, quelles que soient les contraintes d' « *accountability* » imposées à cette autorité.

Ce problème de crédibilité intertemporelle est récurrent dans toute la gestion publique. En matière de politique immobilière, par exemple, il est évidemment souhaitable de conserver des dotations de crédits suffisantes pour assurer l'entretien du parc. Mais il est tentant aussi, au quotidien, d'utiliser ces crédits pour réaliser des actions politiquement plus visibles à l'horizon d'un mandat... d'où la nécessité de trouver des modes d'organisation les protégeant. Donner de l'indépendance aux services en charge de cette mission est assez naturellement une condition nécessaire pour cela, mais non suffisante, car il faut aussi s'assurer qu'ils utilisent bien cette autonomie au profit de l'intérêt général, ce qui suppose une gouvernance efficace.

L'examen d'un cas très ponctuel, – celui des études économiques dans l'administration –, suggère, d'une part, que cette gouvernance est à préciser au cas par cas, d'autre part, que l'évaluation de la performance pour réaliser les missions s'inscrivant dans le long terme est particulièrement importante, à la fois parce que des tâches importantes seraient délaissées sinon, et parce qu'il faut s'assurer que cette indépendance n'a pas été pervertie.

On observe dans ce cas deux types d'études polaires, dont la gouvernance est en effet très différente. Les premières, qui se situent en aval de la statistique, visent à la mettre en perspective, ou à informer le public dans des domaines où le coût des études à réaliser et leur intérêt diffus ne permettent pas d'envisager leur réplique dans un système ouvert (tel que celui qui existe en matière de prévision macroéconomique). Les secondes sont des études de « conseil », pour fournir au gouvernement les éléments d'éclairage sur les réformes qu'il souhaite ou devrait envisager.

Pour les premières, une forte indépendance est justifiée, associée à une exigence de neutralité, le souci devant être de ne chercher ni à aider, ni à nuire au gouvernement en place. Pour cela des règles de déontologie, dans la gestion des publications par exemple, sont nécessaires. Par ailleurs, un élément crucial, qui peut être évalué, est de s'assurer de la qualité scientifique des travaux.

Les études du second type ne réclament pas la même nature d'indépendance, puisqu'il s'agit d'aider le gouvernement, quel qu'il soit. Pour autant leur organisation réclame encore une forme de séparation, pour ne pas délaissier les problèmes structurels qui sont souvent les plus importants, mais dépassent l'horizon d'un gouvernement. Ceci suppose, à nouveau, un contrôle de la qualité scientifique, pour assurer ce qui est fourni par les services, mais aussi un suivi des orientations des études pour garantir, à la fois que des investissements suffisants sont réalisés sur les sujets émergents, mais aussi que les demandes résultant de l'agenda sont servies.

1.2. La mesure de la performance : problèmes de méthodes

L'évaluation des performances du secteur public est ainsi une question réellement délicate. D'un point de vue théorique, elle suppose une réflexion approfondie sur les missions qui lui sont assignées, sur les critères qui doivent être retenus et sur les motivations des agents de l'État.

1.2.1. Spécificité des missions

La notion de mission a été définie par Wilson (1989), dans son étude de la bureaucratie, comme une culture partagée par l'ensemble des membres d'une organisation⁽¹⁾. Cette notion vague a de nombreuses implications sur le fonctionnement du secteur public.

Les objectifs de l'action publique sont complexes, multidimensionnels et souvent flous, notamment quant aux pondérations des différents critères. Au niveau des programmes, cette multiplicité des dimensions de la politique publique se traduit par une multiplicité de tâches. Une éducation de qualité implique l'acquisition de connaissances mesurées par des résultats à des tests standardisés ou à des concours mais aussi l'apprentissage de la citoyenneté et de la vie en société ou l'encouragement de l'esprit de créativité et de curiosité. Elle ne s'apprécie pas de la même manière selon que l'objectif assigné est l'instruction ou la réussite de tous. L'Université a des fonctions de délivrance des diplômes de l'enseignement supérieur mais également de recherche, les professeurs devant aussi participer à sa gestion. L'activité de la police peut s'évaluer par des indicateurs reposant sur le nombre de contraventions, sur le nombre de points retirés sur les permis de conduire, sur le taux d'élucidation des enquêtes ou sur le nombre de délits mais comporte aussi une dimension préventive qui ne s'évalue pas à court terme.

Cette complexité des services publics et l'imprécision des objectifs à tous les niveaux de l'administration distinguent la performance publique de la performance privée. D'une part, les indicateurs pertinents sont souvent imparfaits voire inexistant, la production publique restant difficile à mesurer autrement que par son coût, et non par son « *output* » et *a fortiori* son « *outcome* », dans de nombreux domaines. D'autre part, les résultats de l'action publique dépendent de variables exogènes qui biaisent les évaluations. Ainsi l'état de santé d'une population, imparfaitement mesurable par des indicateurs d'espérance de vie, de mortalité ou de morbidité, dépend de la qualité des soins mais aussi des modes de vie, de l'alimentation, de l'hygiène ou des conditions d'habitat. De même, l'activité d'un commissariat ou d'une école ne peut se mesurer de la même manière dans une banlieue défavorisée ou dans un quartier riche. Par ailleurs, les agences gouvernementales ont moins d'autonomie pour déterminer leur action que les entreprises privées. Elles agissent dans le cadre d'une contrainte budgétaire avec

(1) Cf. Beesley et Gathak (2003) ou Dewatripont, Jewitt et Tirole (1999) pour un exposé.

des dotations de facteurs fixes et des possibilités limitées de substitution de ceux-ci. Au niveau décentralisé, la performance d'un agent ou d'un service est limitée par ces contraintes.

1.2.2. Multiplicité des tâches et évaluation

La multiplicité des dimensions de l'action publique confère un caractère de multiplicité des tâches à l'activité quotidienne des agents de l'État⁽²⁾. Or certaines des tâches sont mesurables (par exemple par les coûts *ex post*), d'autres le sont imparfaitement et d'autres encore sont non mesurables (comme certaines caractéristiques qualitatives). Cela est à l'origine d'une tension entre ces différentes dimensions. Ainsi la mise en œuvre d'une tarification à l'activité des hôpitaux est conçue comme un contrat à prix fixe qui engendre une minimisation des coûts. Or celle-ci peut être obtenue par l'élimination des gaspillages de toute nature et par la recherche de l'efficacité productive mais aussi par une réduction de la qualité des soins fournis ou par une sélection des patients, l'hôpital refusant de traiter les cas les plus coûteux. La sélection des patients ou la qualité des soins étant imparfaitement mesurée, l'hôpital peut être incité à concentrer ses efforts sur la réduction des dépenses au détriment des efforts d'amélioration de la qualité ou d'accueil des patients les plus gravement atteints. Comme le montrent Holmström et Milgrom (1991), mettre des incitations fortes sur une dimension de l'activité (ici l'efficacité productive) implique des effets sur les activités complémentaires ou substituables quand des incitations aussi fortes ne peuvent être mises en œuvre sur celles-ci. Cette problématique se retrouve dans l'ensemble de la fonction publique. Fonder les avancements des professeurs d'Université sur des critères stricts de publications peut les inciter à délaisser leur activité d'enseignement ou leur participation à la gestion des établissements. La répartition des effets sur les différentes tâches peut aussi dépendre des incitations différentes concernant chacune des tâches, dans le domaine de l'enseignement primaire ou secondaire, de la sécurité intérieure ou de l'administration fiscale.

Deux exemples peuvent aider à faire comprendre la réalité et la généralité du problème :

- la gestion des risques et de la sécurité, qui sous des formes diverses, se trouve au cœur de beaucoup d'interventions publiques, et vis-à-vis desquelles le public attend beaucoup de l'État. Celle-ci comporte (au moins) une double dimension de prévention et de gestion des crises, les deux pouvant être contradictoires. Si un responsable est en effet jugé exclusivement sur la première, donc sa capacité à minimiser le nombre d'incidents ou d'accidents, il sera très fortement incité à retarder leur déclaration, ce qui peut s'avérer néfaste par le rapport à la seconde dimension, qui réclame en effet de pouvoir s'organiser au plus vite, tant qu'il est temps...

(2) Cf. Holmström et Milgrom (1991).

- la gestion budgétaire. Dans ce cas, il y a conflit entre la capacité à tenir son « exécution », ou à limiter les besoins immédiats de crédits, et l'intérêt d'identifier tôt des dépenses inéluctables, pour les provisionner. La mesure de la performance des responsables concernés doit en tenir compte.

Les difficultés d'évaluation des performances dans ce contexte de tâches multiples sont renforcées par les nombreuses asymétries d'information qui caractérisent l'activité du secteur public, dès lors que celui-ci est privé de l'information transmise par les marchés. La performance, dans toutes les composantes, dépend des efforts non observables et non mesurables effectués (aléa moral) mais aussi de caractéristiques exogènes également non mesurables (anti-sélection) telles que la compétence des agents ou la productivité. Cette conjugaison d'éléments d'aléa moral et d'anti-sélection renforce la complexité de l'évaluation. Elle peut engendrer des manipulations de l'information transmise lorsque celle-ci est non vérifiable. Elle peut être à l'origine de rentes informationnelles variables selon les modes de financement. Ainsi, si la gravité des pathologies est observable par les offreurs de soins, la tarification à l'activité aura, certes, des effets positifs en termes d'efficacité productive mais conduira à des rentes élevées lorsque l'hôpital accueillera des patients à coûts bas si le prix est fixé à un niveau élevé pour éviter la sélection. Le mode de paiement (c'est-à-dire le mode de financement des services publics) doit donc reposer sur un arbitrage entre l'extraction de la rente et l'efficacité. Un tel arbitrage doit être considéré dans un grand nombre de domaines de l'action publique.

Le secteur de la santé fournit, par ailleurs, à propos des mécanismes de responsabilisation des patients, une illustration de la distinction à opérer entre incitations fortes et niveaux d'incitations « intermédiaires ». Ceux-ci rencontrent en effet le même type d'arbitrage à réaliser entre équité et incitations. À cet égard, l'impact des modalités d'assurance des malades sur la demande de soins est aujourd'hui bien cerné, notamment grâce au *Health Insurance Experiment* (Manning et al., 1987). Globalement, la gratuité des soins entraîne un niveau de consommation supérieur de 46 % à celui que l'on observe lorsque les frais sont intégralement supportés par les ménages. La moitié de cet écart est obtenue en passant de la gratuité totale au ticket modérateur (d'ordre public) de 25 %. L'essentiel de la différence entre ces montants de dépenses résulte de la fréquence des séances de médecin, puis, dans une moindre mesure, de celle des hospitalisations. Par contre, on n'observe pas de différence dans les montants de dépenses par visite ou par admission.

Étant admis que la couverture totale des soins médicaux peut induire un supplément de dépenses de près de 50 % par rapport aux formules comportant un ticket modérateur significatif, la principale question soulevée par les professionnels de la santé face aux résultats du *Health Insurance Experiment* a bien évidemment été celle de l'impact de cette variation de dépenses en termes de santé : les patients couverts à 100 % reçoivent-ils trop de soins, ou ceux soumis à un fort taux de coassurance pas assez ? L'étude complémentaire qui a été réalisée pour répondre à cette question ne

met pas en évidence d'amélioration significative des indicateurs de santé du fait d'un taux de couverture de 100 %, pour les personnes de caractéristiques moyennes.

Le principal impact notable concerne la correction de la vision, l'écart sur la tension artérielle étant la limite de la significativité. Par contre, ces effets apparaissent amplifiés pour les personnes cumulant un faible revenu et un handicap initial important en ces domaines. Ces résultats ne constituent donc pas un argument fort pour exclure un taux de participation des assurés plus élevé à la couverture des dépenses maladie, car les écarts en termes d'indicateurs de santé concernent des domaines (tension, myopie, soins dentaires), où le diagnostic est peu coûteux. L'essentiel de ceux-ci pourrait donc être obtenu à plus faible coût par des programmes spécifiques, plutôt que par une couverture totale de tous les soins.

Outre le fait que les agents réagissent aux incitations financières, cet exemple montre ainsi : que les mécanismes d'incitations sont déterminants pour limiter l'impact des asymétries d'information ; qu'en présence de conflit aigu entre ici l'objectif de responsabilisation du patient, d'une part, et celui de limiter le risque qu'il supporte d'autre part, un niveau incitatif intermédiaire peut avoir beaucoup de valeur ; que la conception des mécanismes d'incitations correspondants doit tenir compte de variables sur lesquelles l'agent a effectivement prise ; et qu'il convient de toujours identifier les dimensions sur lesquelles l'impact d'un mécanisme donné peut être défavorable, de manière à rechercher les instruments d'action complémentaires pour contrôler ces objectifs.

1.2.3. Multiplicités des mandants

Toute organisation repose sur un réseau de relations de type principal-agent ou mandant-mandataire (le « principal » étant, dans le langage de la théorie économique, celui qui est responsable de la définition des missions). L'activité du secteur public implique fondamentalement qu'un agent dépend de plusieurs mandants⁽³⁾ en raison des externalités associées à sa production. Les entreprises publiques doivent répondre à leur ministère de tutelle mais aussi au ministère des finances. Un enseignant doit rendre des comptes à son principal, à l'inspection d'académie mais aussi aux parents d'élèves. Un médecin hospitalier est impliqué dans une relation où la direction de l'hôpital et les patients sont des mandants.

Cette question apparaît essentielle pour les services fournis directement à la population et pour lesquels les citoyens interviennent directement comme des consommateurs. Faire jouer aux usagers un rôle en associant leur satisfaction aux règles d'évaluation et d'incitation peut accroître la performance à condition de bien analyser la structure des relations de type principal-agent caractéristique de l'activité concernée.

Les critères dévaluation de ces principaux étant différents, tout mécanisme d'incitation qui ne serait pas conçu conjointement est nécessairement

(3) Cf. Dixit (2002), Besley et Ghatak (2003) et Tirole (1994).

d'une faible efficacité dans un contexte de tâches multiples. Si les tâches privilégiées par chaque principal sont complémentaires, un problème de passager clandestin peut apparaître, chacun comptant sur l'autre pour mettre en œuvre les incitations adéquates. Si elles sont substituables, une dilution des incitations réduira leur efficacité⁽⁴⁾.

L'illustration de ces principes ne peut être ici qu'heuristique, sachant que, souvent, il y a un *mix* de complémentarités et de substitutions entre ces principaux, selon les dimensions considérées. Par ailleurs, il faut examiner évidemment la manière dont les intérêts de l'agent sont plus ou moins spontanément en ligne avec les différents objectifs de ces mandants.

Le cas de la tutelle des établissements publics et celui de l'organisation des services extérieurs de l'État permettent d'illustrer la pertinence de l'approche. Il est généralement considéré, en effet, que la multiplicité des tutelles constitue un handicap pour la bonne gouvernance des établissements. Mais, si celle-ci traduit la multiplicité des parties prenantes, elle est largement incontournable, d'où la nécessité de chercher les solutions pour la gérer au mieux.

Le problème de passager clandestin évoqué ci-dessus se rencontre dans les établissements dont le pilotage repose sur un partenariat entre différents niveaux d'autorités, locales et centrales, ou ceux associant les partenaires sociaux, comme c'est le cas des grandes assurances publiques. Cette configuration peut contribuer en effet à expliquer les difficultés rencontrées pour y mettre en place des mécanismes de contrôle et de gestion des risques appropriés. Le cas de services extérieurs, qui relèvent à la fois d'une autorité territoriale (préfet ou ambassadeur) et d'autorités sectorielles centrales peut aussi conduire à ce type de problème, qui nourrit, en pratique, les missions communes des corps d'inspection. Mais le phénomène de « passager clandestin » apparaît bien comme quelque chose de robuste (comme le suggère la théorie), quand il s'agit de passer aux « suites » de ces rapports, ou quand il s'agit d'animer, dans la durée, des lieux de coordination des administrateurs de l'État pour les établissements publics. D'où la nécessité d'organiser systématiquement la coopération des « principaux » concernés...

Les cas de dilution des incitations apparaissent lorsqu'un service doit rendre compte à plusieurs tutelles sur des objectifs plus complexes tels que : maîtriser les risques industriels ou sanitaires, tout en favorisant la compétitivité économique (industrie, économie, environnement), ou poursuivre une activité scientifique et universitaire, avec la présence et l'implication dans un service hospitalier, par exemple (santé, université). Ce type de situation soulève, au sens propre, des problèmes de gouvernance, qui dépassent le champ de ce rapport. Mais ceux-ci ont aussi des implications directes pour l'évaluation de la performance. À cet égard, les deux recommandations que l'on pourrait suggérer sont les suivantes :

- s'astreindre dans ce cas à un effort particulier de formulation des missions de manière à ce que l'agent ait une vision claire de celle-ci, et des

(4) Cf. Besley et Ghatak (2003).

conditions dans lesquelles il devrait aborder les éventuelles situations de conflits d'intérêt ;

- s'inspirer de ce qu'ont fait les entreprises en développant des rapports particuliers, par exemple pour les aspects sociaux ou le développement durable, pour rendre compte à leurs différentes parties prenantes, étant entendu que leur crédibilité d'ensemble doit évidemment être assurée.

2. Incitations et gestion publique

« Une autre interprétation de la théorie de Rousseau, est que le gouvernement est seulement composé d'employés qui mettent en œuvre les politiques décidées par la « volonté du peuple ». Leurs motivations privées s'expliqueraient simplement dans ce cadre : ils obéissent avec précision aux ordres du peuple pour conserver leur emploi, car la moindre désobéissance signifierait leur renvoi immédiat. Comme le montre l'ensemble de notre étude, cette vision est incompatible avec l'incertitude et la division du travail ».

Downs (1957), cité par J-J. Laffont (2000).

L'analyse des problèmes de méthode rencontrés pour l'élaboration d'indicateurs de performance nous a conduit à établir un lien entre la mesure de la performance et la qualité des incitations. On a signalé, à cet égard, certaines spécificités de la gestion publique, concernant ses missions ou la multiplicité de ses tâches et mandants, et suggéré que l'on disposait de concepts solides pour les aborder, mais en restant toutefois au stade de l'illustration. Il convient maintenant, pour progresser, de préciser ce lien, à partir d'une approche générale de ces problèmes d'incitations.

2.1. La vision économique de l'intervention publique aujourd'hui

2.1.1. Objectifs, motivations des agents publics et contraintes statutaires

Les agents publics, quels que soient leur grade et leur fonction, ont des objectifs personnels qui peuvent différer de l'intérêt général. Certes, on peut concevoir que ceux qui ont choisi de servir l'État ont une fonction d'utilité différente de ceux qui travaillent dans le secteur privé et qu'ils attachent plus d'importance au bien-être social et au service de la population. Mais on peut aussi penser que le choix d'une carrière sans risque de perte d'emploi et comportant une garantie de stabilité du revenu révèle une aversion au risque plus forte que dans le secteur privé. Par ailleurs, on ne peut pas non plus écarter le fait que la faiblesse relative des incitations dans le secteur public par rapport au secteur privé influence le choix d'une carrière au sens où ceux qui ont une désutilité de l'effort plus forte la choisirait de manière à devoir fournir un effort moindre que dans le secteur privé. Or les agents publics sont hétérogènes et leurs préférences individuelles n'impliquent pas

que tous aient le même intérêt pour le bien-être social ni la même désutilité de l'effort. Les plus altruistes n'ont pas besoin d'être incités à agir dans le sens de l'intérêt général, leur motivation personnelle les amenant d'eux-mêmes à faire l'effort le plus important au service de la population. En revanche, les moins altruistes ne le feront que s'ils sont induits à le faire par des règles incitatives de rémunération ou d'avancement.

Les agents publics ont, d'autre part, des préférences personnelles concernant les tâches qu'ils doivent effectuer. Un professeur d'Université peut préférer la recherche à l'enseignement ou l'inverse. Un militaire de carrière peut avoir plus d'intérêt dans l'une ou l'autre des missions de l'armée. Certains privilégieront les tâches qui les mettent en contact avec le public, d'autres non.

Compte tenu de ces caractéristiques des préférences des agents de l'État, leurs performances – en termes de coût ou de qualité du service – varieront en fonction du poids qu'ils attachent au service du public, de l'intérêt que présentent pour eux les différentes tâches et de l'importance de leur désutilité de l'effort. Par suite, elles seront conditionnées par les incitations mises en place pour atteindre les objectifs de l'État. Le coût et la qualité des services sont donc endogènes. Cette idée que la mise en place de la LOLF devrait contribuer à mettre en évidence s'oppose aux pratiques budgétaires traditionnelles qui consistent à raisonner dans le cadre strict d'une enveloppe financière donnée et d'une fonction de production rigide. Ces contraintes, qui sont des freins à l'efficacité, doivent être prises en compte dans tout mécanisme d'évaluation des performances. De même, les agents publics exercent leur activité dans un cadre statutaire qui peut nuire à la mise en œuvre d'incitations efficaces et qui doit être pris en compte dans toute analyse, qu'il s'agisse de l'emploi à vie, des règles d'avancement à l'ancienneté ou de l'existence d'une grille indiciaire. D'autres contraintes, coutumières ou quasi statutaires, viennent, de plus, limiter la force des incitations susceptibles d'être instaurées. Ainsi la rigidité du marché interne du travail se traduit par des limites à la mobilité et au redéploiement des moyens. De même, la lourdeur des procédures administratives, qu'elles concernent les publications d'emplois vacants ou les achats d'équipement, constitue un obstacle évident aux ajustements de la fonction de production des services publics rendus nécessaires par le progrès technique. Alors que des substitutions de facteurs s'effectuent de manière relativement rapide dans le secteur privé, elles sont lentes, voire inexistantes à court terme, dans le secteur public. La non-réactivité à l'innovation constitue, à cet égard, une composante non négligeable de la faiblesse des performances de certains services publics. Les évolutions comparatives de l'hospitalisation privée et publique illustrent ce phénomène dans de nombreux domaines, notamment dans celui de la chirurgie ambulatoire (*cf.* le complément de Bousquet et Coulomb).

Ces contraintes, juridiques et administratives sont, à l'évidence, des freins à l'efficacité dans la mesure où elles sont de nature à limiter les possibilités d'action d'agents performants. D'autres contraintes, d'ordre informationnel, viennent restreindre la mise en œuvre des politiques publiques en empêchant la distinction entre agents efficaces et agents inefficaces. Elles nécessitent l'instauration de mécanismes d'incitation pour en limiter les effets négatifs.

La reconnaissance de ces problèmes d'incitations dans la gestion publique a été progressive. Elle a modifié profondément la vision des économistes sur l'intervention publique, à la fois quant à son champ et à ses modalités d'intervention, et quant à l'attention à porter aux questions de performance des organisations. Il apparaît nécessaire de s'arrêter un temps sur le premier aspect, pour bien comprendre les problèmes associés à la mesure de la performance.

2.1.2. Fondements et modalités de l'intervention

Pour mettre en évidence les spécificités de la gestion publique, rappelons qu'il convient, en premier lieu, de définir le périmètre de l'intervention publique puis de caractériser les objectifs – les missions – de celle-ci. Il faut ensuite s'interroger sur la structure des organisations publiques, sur leurs performances et leurs motivations.

Il est traditionnel de définir le rôle de l'État par référence aux défaillances du marché, les services publics étant associés à des biens et à des services qu'un marché non réglementé ne pourrait produire ou produirait en quantité insuffisante du point de vue de l'efficacité ou de l'équité. Cette définition renvoie à trois raisons principales de l'intervention publique (encadré 1).

1. Principales raisons de l'intervention publique

1. La présence d'externalités

Cette première justification de l'action de l'État concerne, d'abord, les biens publics purs, c'est-à-dire ceux qui possèdent la propriété de non-exclusion, la consommation par un citoyen ne réduisant pas celle des autres.

Lorsque cette propriété est vérifiée au niveau de la nation, les biens publics correspondent aux fonctions régaliennes traditionnelles de l'État : défense, sécurité, justice, administration générale. Entre ces biens publics purs et les biens privés, il existe un grand nombre de biens et services générateurs de bénéfices collectifs et justifiant une intervention publique : les biens publics locaux ou les biens dont la consommation ou la production s'accompagne d'effets externes non médiatisés par le système de prix. Cela concerne aussi bien les interactions hors marché négatives (comme les atteintes à l'environnement) que celles qui impliquent des effets externes positifs. Ainsi l'amélioration de l'état de santé ou du niveau d'éducation de chaque citoyen bénéficie à l'ensemble de la collectivité.

Pour ce type de biens et services, l'intervention de l'État est justifiée par l'incapacité du marché à intégrer dans les prix les effets externes, la production privée étant socialement insuffisante (voire nulle) dans le cas d'externalités positives et excédentaires dans le cas d'externalités négatives. Cependant, précisons d'ores et déjà qu'intervention publique n'implique pas nécessairement production publique et qu'un grand nombre d'arrangements institutionnels sont concevables. Dans tous les cas, elle suppose cependant la capacité d'évaluer les coûts des dommages environnementaux ou la valeur des aménités environnementales.

Des méthodes ont été développées pour cela, et on a assisté plus généralement à un renouveau des méthodes coûts-avantages dans la plupart des domaines de l'intervention publique compris pour évaluer les réglementations (cf. Cahiers français, 2005).

2. Les imperfections des marchés

S'il est reconnu qu'un marché parfait peut conduire à la meilleure allocation des ressources pour la collectivité, deux catégories d'imperfections peuvent se traduire par des inefficacités sociales et par une production socialement insuffisante.

En premier lieu, la production de certains biens ne peut s'effectuer que dans des conditions technologiques qui se traduisent par l'émergence de monopoles. Cette situation de monopole naturel résulte de la sous additivité des coûts et est associée à l'importance des charges fixes. La fourniture publique ou la régulation de ces activités permettent de réduire les distorsions et les rentes induites par un monopole privé. Cette situation est cependant de nature à évoluer dans le temps en fonction des technologies (ainsi la caractéristique de monopole naturel des services de télécommunications s'est réduite avec l'apparition de la téléphonie sans fil) et de la possibilité de séparer la gestion de l'activité de celle de l'infrastructure. À nouveau, différents arrangements institutionnels sont concevables pour mettre en œuvre ce mode d'intervention publique.

En second lieu, les demandes transmises au marché peuvent ne pas correspondre aux besoins véritables des agents. Ainsi, les individus peuvent mal apprécier les bénéfices résultant de leur consommation de services de santé et d'éducation ou les effets néfastes des consommations de certains biens (tabac, drogue, alcool) ou de certains comportements (vitesse excessive, obésité). La mise sous tutelle de certains biens (*merit goods*) se traduit ainsi par une intervention de l'État qui se substitue aux agents pour déterminer le niveau de leur consommation (éducation obligatoire, vaccination obligatoire, interdictions diverses).

3. Équité et redistribution

L'allocation des ressources par le marché se traduit par une répartition des revenus qui reflète à la fois la répartition des compétences et des talents et le pouvoir de marché. Elle n'a aucune raison d'être équitable ou juste, quel que soit le sens que l'on donne à ces termes. Un gouvernement peut juger que certaines inégalités sont incompatibles avec une société civilisée ou peuvent engendrer des crises sociales. Le second théorème de l'économie du bien-être nous enseigne qu'une allocation optimale particulière peut être réalisée dans une économie de propriété privée par un système de prix concurrentiels et une redistribution forfaitaire des dotations individuelles. L'impossibilité d'effectuer celle-ci conduit à des politiques redistributives de second rang (donc socialement coûteuses en termes d'efficacité). La redistribution constitue donc une troisième justification de l'intervention de l'État qui peut mettre en œuvre des transferts financiers ou fournir gratuitement certains biens (notamment ceux qui sont liés à l'éducation et la santé) à la population et ce pour des motifs d'équité. Les missions de service public peuvent ainsi se traduire par des obligations de non-exclusion des plus défavorisés (cf. à ce sujet Cohen et Henry, 1997) ou d'égalité d'accès aux services fournis par l'État. De même, l'existence de biens sous tutelle correspond à une préférence pour l'équité concernant la consommation de certains biens (au sens de l'égalitarisme spécifique de Tobin, 1970).

Ces fondements de l'intervention publique sont peu discutables, et ils demeurent inchangés par rapport à la vision qui s'était constituée après-guerre. Cependant, leur mise en œuvre peut se concevoir dans différents types d'organisation, impliquant des frontières différentes entre public et privé, et ayant des implications variables en termes d'incitations. Précisons, à ce sujet, le vocabulaire en reprenant les distinctions traditionnelles entre services publics et secteur public (cf. Grout et Stevens, 2003). Un « service public » est un service fourni à un grand nombre de citoyens et pour lequel l'implication de l'État est fondée sur l'une ou l'autre des raisons exposées plus haut. Le « secteur public » comprend les activités économiques contrôlées par le gouvernement, la plupart d'entre elles étant des services publics. Une question centrale pour toute réforme de l'État est donc celle du rôle du secteur public et du secteur privé dans la fourniture des services publics. Quelles activités doivent-elles relever de la fourniture publique ? Cette question fondamentale renvoie à celle des performances et des incitations à mettre en œuvre pour atteindre les objectifs des politiques publiques. Elle suppose une réflexion sur les structures organisationnelles, sur la répartition de l'information aux différents niveaux du processus de décision, sur les motivations des agents publics et sur la nature des contrats que l'État peut passer avec les administrations décentralisées et les agents. Cette question est au cœur de ce rapport.

Sans anticiper sur les conclusions, notons, d'abord, que les frontières du secteur public sont mouvantes, qu'elles évoluent en fonction des directives européennes, des changements technologiques et des choix idéologiques. Notons ainsi que le débat « fourniture publique contre fourniture privée » de services publics ne se poserait pas si des contrats complets contingents à tous les états de la nature pouvaient être signés. La contrainte qui résulte de l'incomplétude des contrats et de l'asymétrie d'information qui l'accompagne joue un rôle déterminant quant au choix du type d'organisation. Par ailleurs, le caractère multidimensionnel des missions de service public se traduit par l'existence de critères d'évaluation qualitatifs que l'on ne peut spécifier dans un contrat et donc de dimensions de l'action publique pour lesquelles la performance sera mal mesurée et susceptible de manipulations. Cette difficulté semble essentielle dans la mise en œuvre de la LOLF.

La multitude des arrangements institutionnels est bien illustrée par l'article de Grout et Stevens (2003). Elle se retrouve dans les comparaisons internationales.

Ainsi, l'assurance-maladie est un bien dont la fourniture justifie une intervention publique tant pour des raisons d'efficacité des marchés, en raison des asymétries d'information, que pour des raisons d'équité ou de mise sous tutelle (cf. Mougeot, 1999). Or cette assurance est fournie par l'État dans certains pays, par des marchés réglementés dans d'autres (cf. encadré 2). Il en est de même pour la production de soins hospitaliers, de services d'éducation ou de transport. Lorsque la fourniture publique coexiste avec une fourniture privée, *une concurrence par comparaison* peut être instaurée,

2. Diversité des modalités d'intervention de l'État

Cette diversité peut-être illustrée premièrement à propos des processus de libéralisation des entreprises de réseau.

En effet, la libéralisation ne met pas en cause le maintien d'obligation de service public, qui peut être assurée selon des modalités qui diffèrent selon les secteurs : mise aux enchères éventuelle des obligations, définition d'un service universel financé par un fonds spécifique comme celui des télécommunications, définition d'un service universel et de services réservés permettant de le financer dans le secteur postal. Dans tous les cas, l'ouverture à la concurrence limite cependant les possibilités de subventions croisées entre activités, et incite à la transparence dans la définition de ces missions, et dans les conditions de leur financement.

Des subventions croisées restent néanmoins possibles par le biais des charges d'accès versées aux activités demeurant en situation de monopole, si celles-ci ne peuvent être contournées. Par contre, il va de soi que l'ouverture à la concurrence oblige en général à supprimer les subventions croisées, compte tenu du risque d'écroulement. À noter que le choix de l'entreprise en charge de réaliser les missions de service public peut éventuellement être réalisé par mise en concurrence. La combinaison des critères sur le financement (impôt, ou péréquation, et champ de la péréquation), sur l'extension des droits exclusifs et sur les conditions de l'offre de ces services permet ainsi une large gamme de solutions. Celles-ci vont en effet de l'extension du secteur réservé à la création de fonds financés par l'ensemble des entreprises présentes sur le marché, dont le produit est versé à l'entreprise en charge des missions d'intérêt général, opérateur historique ou entreprise, choisie par appel d'offres. Tendanciellement, ce recours à des appels d'offres s'est développé.

De même, dans le secteur de la santé, différents pays ont imaginé des institutions originales pour combiner recours aux marchés d'assurance et principe de couverture universelle de la population.

Pour cela, la Suisse oblige à proposer un contrat de base à des conditions uniformes, l'assureur ne pouvant refuser de couvrir quelqu'un. Les Pays-Bas, eux, opèrent une séparation entre la collecte des primes et l'assurance, les compagnies recevant une dotation par personne assurée ajustée au risque représenté par sa clientèle.

Le secteur de l'environnement fournit, enfin, avec le développement des écotaxes et des marchés de permis d'émissions, une autre illustration de cette diversification des instruments d'intervention publique. L'enjeu, dans ce cas, est la mise en place d'instruments de prévention minimisant les coûts de protection à engager pour atteindre les objectifs de protection visés, contrairement aux approches traditionnelles, qui cherchent à les atteindre par la réglementation.

ce qui facilite l'évaluation des performances et la mise en œuvre de mécanismes incitatifs efficaces dans une situation d'aléa moral ou, plus généralement, lorsque la corrélation des informations privées rend possible le recours à des règles de paiement prenant en compte cette corrélation et donc

conduisant à l'élimination des rentes informationnelles. La comparaison des indicateurs de performances de différentes compagnies assurant le transport de passagers entre la Corse et le continent, de cliniques privées et d'hôpitaux fournissant le même service identifié par un groupe homogène de maladies (GHM) sont de nature à améliorer sérieusement l'efficacité du secteur public. En revanche, dans de nombreux domaines d'activité, le secteur public occupe une position de monopole qui rend difficile l'évaluation de la performance faute de références externes. Cependant, des comparaisons internes au secteur public restent envisageables.

Lorsque différents modes d'organisation sont concevables, leur choix doit être guidé en fonction des objectifs de l'État, des motivations des agents et des contraintes contractuelles et informationnelles. Considérons maintenant la délicate question de la définition des missions.

2.1.3. Objectifs de l'action publique et missions

L'analyse normative « traditionnelle » de l'économie publique supposait que le gouvernement, c'est-à-dire l'ensemble des hommes politiques qui contrôlent l'appareil d'État et déterminent les politiques publiques, recherche l'intérêt général (ou maximise le bien-être social). Cette hypothèse du gouvernement bienveillant se traduit par des objectifs de surplus collectif qui expriment les arbitrages de l'État entre l'efficacité et l'équité. Laffont (2000) a bien exposé le travail de l'économiste lorsque le gouvernement est bienveillant et bien informé. Il a aussi montré les limites de cette hypothèse.

D'un point de vue positif, le pouvoir est exercé par des hommes politiques qui, certes, peuvent être motivés par la recherche du bien-être social mais sont aussi soumis à un processus électoral et aux pressions de groupes d'intérêt. La contradiction entre les principes moraux et les combats politiques avait déjà été soulignée par Machiavel. L'économie politique contemporaine décrit les politiques résultant des interactions avec les groupes de pression et avec les partis politiques (*cf.* Laffont, 2000 et 1999). Les gouvernements sont ainsi amenés à prendre des décisions ou à poursuivre des objectifs susceptibles de leur rallier une majorité dans les élections, ce qui conduit, dans certaines configurations, à privilégier les préférences de l'électeur médian. Ils peuvent faire l'objet de tentatives de captation de la part de groupes de pression – internes ou externes à l'administration – pour qu'ils orientent la politique en leur faveur.

Par suite, les gouvernements n'optimisent qu'imparfaitement le bien être social, ce qui a des conséquences importantes sur la gestion publique. Comme le rappelle Laffont (1999), celle-ci domine clairement la gestion privée du point de vue de l'intérêt général lorsque l'État est bienveillant et les contrats complets⁽⁵⁾. Il n'en est plus de même lorsque les contrats sont incomplets et lorsque le processus démocratique laisse une marge de discrétion aux hom-

(5) *Cf.* également Sappington et Stiglitz (1987).

mes politiques et donc une place pour l'action des groupes de pression. Ajoutons à cela que le processus politique qui conduit à la définition des objectifs des politiques de l'État est soumis aux variations des préférences des électeurs et implique un manque de cohérence intertemporelle qui réduit les possibilités d'engagement à long terme⁽⁶⁾.

Les conséquences de cette possibilité de capture de décideurs publics par des groupes de pression sur la réforme de l'état ont été analysées par le rapport Laffont (1999). On peut cependant s'interroger sur son influence sur la définition des missions de l'État. En premier lieu, le cadre institutionnel dans lequel les politiques sont menées est lui-même le résultat de politiques qui ont pu refléter les choix des groupes de pression. Les institutions universitaires ou celles qui régissent le système de santé en sont de bons exemples. En second lieu, si la finalité de la recherche de l'intérêt général est, sans conteste, préférable à celle de la recherche du profit, les finalités concrètes mêlant l'intérêt général et les objectifs privés des hommes politiques sont plus complexes à apprécier. D'une part, elles sont, par nature, multidimensionnelles, l'intervention publique ne pouvant s'évaluer par un seul critère alors que l'activité du secteur privé peut l'être par le seul critère du profit⁽⁷⁾. D'autre part, les poids associés aux différentes dimensions de l'action publique peuvent refléter le poids des groupes de pression et sont, en général, inconnus. Ainsi l'arbitrage entre la qualité des soins et le coût pour la collectivité de la fourniture de ceux-ci peut dépendre de l'influence politique des professions médicales qui plaident pour une hausse de la dépense, cet argument de maximisation du revenu se traduisant aisément en revendication d'amélioration de la qualité.

2.2. Contraintes informationnelles, rentes et incitations

2.2.1. Relations « principal-agent »

Comme toute organisation, le secteur public est caractérisé par un ensemble de relations de type « principal-agent » dans lesquelles les asymétries d'information jouent un rôle essentiel. Nous avons vu plus haut comment la conjugaison de l'aléa moral, de l'anti-sélection et d'une incertitude exogène rend difficile l'évaluation des agents de l'État et ceci d'autant plus que leur activité s'exerce dans un contexte de tâches multiples. La rente informationnelle résulte du fait que le principal a moins d'information que l'agent sur ce qui est possible ou sur la manière dont les objectifs du principal peuvent être mis en œuvre. L'agent peut donc exploiter cet avantage informationnel pour réduire son effort, que cet effort concerne la ré-

(6) Cf. Tirole (1994).

(7) C'est ce que souligne le rapport Pochard selon qui « le caractère spécifique des services publics implique... que la réalisation des performances soit appréciée par rapport aux objectifs fixés par l'autorité publique et qu'elle intervienne sur la base de tous les critères pertinents et pas seulement de critères purement financiers ».

duction du coût ou l'amélioration de la qualité du service rendu à la population. La fourniture de l'effort impliquant une désutilité pour l'agent, celui-ci s'approprie une rente qui, certes, n'a pas le caractère monétaire qu'a, par exemple, une rente de monopole, mais possède toutes les propriétés d'une rente en ce sens qu'un service identique pourrait être obtenu à un moindre coût ou qu'un service de meilleure qualité pourrait être obtenu au même coût.

La structure hiérarchique de l'administration contribue à l'émergence de relations de ce type à plusieurs niveaux. À chaque échelon, un responsable supervise l'action de ceux qui sont à un niveau hiérarchique inférieur et rend compte de cette activité à un responsable hiérarchiquement supérieur. Une telle structure crée, naturellement, des possibilités de collusion entre les agents et leur superviseur dans le but d'accroître et de partager la rente informationnelle. Comme nous l'avons vu à propos de l'Université ou des hôpitaux publics, des coalitions de cette nature, qui s'exercent au détriment de l'intérêt général et des usagers, sont fréquentes dans le secteur public.

Dans ce contexte, quelles incitations mettre en œuvre pour que les agents agissent en conformité avec les missions du service public ? Pour répondre à cette question, notons, d'abord, que l'évaluation de la performance des agents doit être distinguée de celle des services (au sens large) dans lesquels les agents interviennent et de l'évaluation des politiques publiques. Chaque agent de l'État agit dans un cadre statutaire défini par la loi, en fonction des politiques déterminées par le gouvernement et des directives assignées à son service et à lui-même. L'efficacité de ces agents ne peut s'apprécier que de manière relative en tenant compte de cet environnement. L'efficacité de l'action publique doit, quant à elle, s'apprécier globalement.

Un second élément à prendre en compte résulte du caractère multidimensionnel des tâches qui réduit nécessairement la force des incitations. Lorsqu'il perçoit un salaire fixe, un agent public est peu incité à diminuer le coût de fourniture de son service si l'effort de réduction de ce coût implique une perte d'utilité. Cependant, si la qualité du service n'est pas observable, un mécanisme qui récompenserait ceux qui diminuent la dépense et pénaliserait ceux qui dépensent plus qu'une certaine valeur de référence pourrait conduire à une baisse de la qualité, c'est-à-dire de la composante non observable de l'effort. À l'arbitrage rente-efficacité déjà mis en évidence s'ajoutent ainsi des arbitrages complexes entre les différentes tâches en fonction de leur mesurabilité, ce qui limite la force des incitations et contribue au maintien de certaines rentes.

Pour achever cette mise en perspective de l'analyse économique de la gestion publique, il faut encore tenir compte de deux autres caractéristiques de l'activité. En premier lieu, il faut noter la faiblesse des incitations monétaires dans l'administration et l'importance du rôle des perspectives de carrière. Les « stocks options », l'intéressement aux bénéficiaires et les primes individualisées constituent des modes monétaires d'incitation plus fréquentes dans le secteur privé que dans le secteur public. En revanche, l'effet que la per-

formance d'un fonctionnaire peut avoir ses probabilités de promotion et donc sur sa carrière future joue un rôle essentiel. Cette caractéristique du secteur public implique un raisonnement en plusieurs étapes. En tenant compte de l'effet futur de son effort, un agent public peut être incité à améliorer sa performance même avec un salaire fixe. Comme l'a indiqué Holmström (1982), un schéma d'incitation fondé sur les promotions futures suppose que la performance soit observable par ceux qui décident des promotions futures, que le fait de signaler sa compétence future par un effort aujourd'hui ne soit pas trop élevé. Par ailleurs, la question des incitations fondées sur les perspectives de carrière peut conduire à un effort d'autant plus grand que l'incertitude sur les aptitudes est élevée, les agents devant convaincre leur supérieur de leur qualité⁽⁸⁾. Cependant, la force des incitations liées aux promotions se heurte au caractère multidimensionnel des tâches et à la complexité des missions. Comme l'ont montré Dewatripont, Jewitt et Tirole (1999), l'accroissement du nombre des tâches assignées à un agent réduit son effort total car le lien entre la performance et l'évaluation de la compétence par le supérieur hiérarchique devient trop lâche. Cette observation implique une corrélation entre la manière dont une organisation ou un agent doit rendre des comptes et la spécificité de ses objectifs. Par ailleurs, un nouvel arbitrage peut être mis en évidence entre les variations de la performance globale et l'effort, les agents effectuant un effort pour convaincre de leur compétence d'autant plus important que l'incertitude est grande. Le principal peut avoir intérêt à spécialiser les agents dans un sous-ensemble de tâches pour lesquelles il y a une corrélation entre l'effort et la compétence, ce qui s'effectue au détriment des autres tâches, et donc de la performance globale.

Enfin, il convient de tenir compte de la nécessité de prendre en considération le travail en équipe dans la définition des incitations des services eux-mêmes (*cf.* Holmstrom, 1982). Lorsqu'il existe des complémentarités dans la production, les incitations individuelles conduisent à des comportements de passager clandestin nuisibles aux résultats du groupe. Les incitations des services sont liées à la notion d'évaluation par les pairs. Relier les rémunérations des agents aux performances du groupe peut en effet inciter les membres du groupe à procéder à des évaluations internes, éliminant ainsi les comportements de passager clandestin. Les incitations de groupe favorisant la promotion des externalités positives sont particulièrement opportunes quand la coopération des membres du groupe est un facteur important de réussite. Rémunérer la performance individuelle peut alors accroître le coût de la coopération et donc réduire la performance du groupe⁽⁹⁾.

(8) *Cf.* Dixit (2002).

(9) *Cf.* Milgrom et Roberts (1990).

3. La rémunération au mérite au sein de la fonction publique des pays de l'OCDE^(*)

Les administrations, souvent dominées par des hiérarchies verticales, s'apparentent plutôt à la firme fordienne. Certaines d'entre elles, subissant l'évolution d'ensemble de la société, ont malgré tout recherché une flexibilisation des salaires dès les années cinquante. Cette tentative faite dans la plupart des fonctions publiques des pays de l'OCDE n'a pas eu de suite. La rémunération au mérite (RAM) a peu à peu été détournée de son but initial pour devenir une rémunération à l'ancienneté.

Une remise en question des systèmes de rémunération des fonctionnaires est intervenue à partir des années quatre-vingt, les difficultés budgétaires des États de l'OCDE les incitant à rechercher l'efficacité de la dépense publique dans tous les domaines où cela était possible. Le système de promotion étant en panne dans de nombreux pays à cause de la proportion élevée d'agents de plus de cinquante ans, il n'y avait plus de levier de motivation des agents. La RAM est apparue comme un instrument permettant d'attirer et de retenir les employés dynamiques et performants.

Pratiquement tous les pays ont aujourd'hui un système d'évaluation personnalisée de leurs agents, qu'ils essaient d'utiliser pour gérer la performance en conformité avec les objectifs institutionnels. Un tel système s'est mis en place dès les années quatre-vingt au Danemark, au Pays-Bas, en Nouvelle-Zélande, en Espagne, en Suède, au Royaume-Uni et, dans une moindre mesure, en Finlande et Allemagne. Ces pays ont été suivis au début des années quatre-vingt-dix par l'Australie, l'Irlande et l'Italie et, plus récemment, par l'Allemagne, la Corée, la Suisse et quelques pays d'Europe de l'Est. La France et les États-Unis sont en phase de test depuis 2004.

Les enseignements à retirer de ces expériences étrangères sont les suivants :

- les systèmes de rémunération liée à la performance restent localisés. Ils ne concernent, en général, que certains services publics (ou certaines catégories de personnel). La part de rémunération liée à la performance dans la rémunération totale reste modeste (en moyenne 10 % pour les personnels et entre 20 et 25 % pour les dirigeants) ;
- la RAM est d'autant plus répandue que le pays a une gestion des ressources humaines et des budgets décentralisée. De manière générale, la conception du système est plus décentralisée pour les employés que pour les managers. Mais la délégation que nécessite la RAM pose certains problèmes de cohérence et d'équité entre les différents services, c'est pourquoi des comités de contrôle sont souvent mis en place. Certains pays ont d'ailleurs constitué des bases de données leur permettant de faire des études statistiques pour s'assurer du bon fonctionnement du système ;

(*) Cf. Martine Perbet, 2007.

- certaines tendances communes se dégagent. L'évaluation est devenue moins standardisée, moins quantifiable et se fonde davantage sur un dialogue avec la hiérarchie, voire à partir du regard des pairs ou des collaborateurs. Ces dernières années, on assiste aussi à un glissement vers des objectifs collectifs et donc vers une rémunération à la performance liée aux résultats de l'unité (pallier à la non-coopération). D'autres formes de gratifications se développent, notamment au Royaume-Uni (congés, horaires privilégiés...) et en Australie (reconnaissance officielle, accès à des formations, des conférences ou des séminaires à l'étranger, publication dans des médias internes ou externes...);

- l'évaluation ne doit pas être trop compliquée et il n'est pas nécessaire de demander une différenciation trop détaillée pour les travailleurs dans la moyenne. Il faut surtout distinguer les extrêmes, ce qui est plus facile à faire pour les managers. Cela peut se faire grâce à des quotas mais ceux-ci sont à manier avec attention. En effet, ils sont mal acceptés par les employés et peuvent embarrasser les managers, qui vont alors faire tourner les agents dans les groupes. Dans le cas où l'évaluation de la performance individuelle peut avoir un impact négatif sur le travail d'une équipe, il faut plutôt mettre en place des rémunérations par équipe ;

- faire adhérer les agents au système est capital pour son implémentation. Or ceux-ci, ayant souvent un « emploi à vie », sont spontanément peu enclins à accepter ce nouveau mode de rémunération (aversion au risque). Dans les pays qui se sont engagés dans des négociations collectives, le système a été mieux accepté par les agents et l'application s'est faite progressivement, ce qui a permis de tirer des leçons des expériences au fur et à mesure ;

- la RAM coûte. Elle mobilise des ressources humaines (temps de travail, formation des cadres) et budgétaires car il est nécessaire de donner une compensation financière aux perdants (tout du moins au démarrage). Les dépenses liées à la RAM sont souvent sous-estimées. En cas de budget limité, il vaut mieux que le maximum d'agents soient intégrés au système même si cela a pour conséquence des primes plus faibles car l'effet sur la performance vient beaucoup plus du signal envoyé à l'agent quand on l'intègre à ce système que du pouvoir incitatif des sommes en jeu.

L'intérêt de ces systèmes, est moins d'agir sur la motivation des agents (les perspectives de carrière, la mobilité, l'intérêt du travail sont en général plus motivants) que de créer une culture de l'organisation. La RAM agit essentiellement sur la performance par une rénovation du management. Elle permet de mettre en place de *manière plus crédible la gestion par objectifs*, de donner plus de pouvoir aux managers, d'améliorer le *dialogue employé-manager*. Elle permet souvent de *clarifier les tâches* des postes et de rendre le système plus transparent. Enfin, elle semble aussi avoir un *impact positif sur le recrutement*. L'impact semble donc plus dû aux changements qui accompagnent la RAM qu'à l'effet incitatif des primes ou bonus accordés aux agents méritants. Vis-à-vis de l'extérieur, il s'agit aussi d'améliorer l'image de la fonction publique en montrant que les fonctionnaires doivent aussi répondre de leur travail.

2.2.2. Degré de spécificité du secteur public

Avant d'aborder plus précisément les problèmes d'organisation de l'évaluation, il est utile de revenir sur la question du degré de spécificité du secteur public, par rapport au secteur privé, au regard de ces problèmes d'incitations et de mesure de la performance. Les deux visions polaires habituelles en ce domaine consistent, pour l'une à suggérer, qu'au fond, ces problèmes se rencontrent dans toutes les organisations et donc qu'il « suffit » de transposer au secteur public ce qui se fait dans le privé, et, pour l'autre, à arguer de spécificités faisant du secteur public un type d'organisation à part.

Dans l'ensemble (et les références utilisées ci-dessus vont en ce sens), les problèmes d'incitations rencontrés dans le contexte public ne sont pas de nature spécifique, si bien que la gestion publique a tout intérêt à évaluer ses réformes visant la performance en prenant en compte le retour d'expérience de l'utilisation de mécanismes similaires dans le privé.

Beaucoup de problèmes rencontrés dans la gestion publique présentent en effet de nombreux symptômes communs avec ceux qui retiennent l'attention lorsque l'on analyse le gouvernement des entreprises en général : stratégies d'enracinement des dirigeants, construction « d'empires » ; prise de risque excessive ou insuffisante selon les cas ; dispersion des activités ; productivité insuffisante, etc. L'analyse économique tend à les relier à un problème d'agence et de coûts de transaction : les dirigeants possèdent de l'information « privée » et peuvent ne pas agir dans l'intérêt de l'entreprise, ce qui nécessite de trouver des incitations adaptées pour les responsabiliser. Cette ligne de réflexion conduit à souligner que le gouvernement des agences publiques souffre intrinsèquement de l'absence de mesure objective externe des performances et de ne pouvoir donc compter que sur le *monitoring* interne pour « discipliner » leurs dirigeants. La qualité de l'information sur la performance est donc déterminante, et peut être favorisée par la mise en place d'instruments, transposant au secteur public les démarches du secteur privé.

Cependant, le « il suffit de transposer... » ne traduit pas le fait que la compréhension de ces problèmes d'incitations dans la gestion privée et de leur rôle structurant dans l'organisation des firmes est elle-même récente, l'ouvrage de Tirole (2005) étant fondateur. En tout état de cause, il ne s'agirait donc pas d'adopter des « recettes », mais une démarche de travail générale. Surtout la dimension non marchande de nombreux services fournis et les caractères multitâches-multiprincipaux évoqués ci-dessus exacerbent ces problèmes d'incitations lorsqu'il s'agit du secteur public.

En effet, à l'exception de quelques contributions historiques mais isolées, ce n'est que récemment que l'analyse économique s'est intéressée à ces questions dans le cadre même de la firme. La théorie tendait, dans sa version élémentaire, à n'en retenir que son rôle de centre de décision technique, laissant ouvertes les questions posées par Coase avant-guerre sur les frontières des entreprises, et la délimitation d'un territoire non régi par le marché. De même, le théorème de Modigliani-Miller qui montre que, dans un cadre simplifié, la valeur de l'entreprise ne dépend pas d'un ratio endet-

tement-financement par actions, a limité singulièrement les possibilités de dialogue entre économistes et gestionnaires.

En liaison avec un certain nombre de questions d'actualité, telles le développement des marchés financiers, le rôle des OPA, le rôle des fonds de pension, des banques, des commissaires aux comptes, du conseil d'administration, l'approche économique de l'entreprise a fortement progressé, à la fois d'un point de vue théorique et d'un point de vue empirique. Elle reste cependant en pleine évolution : certains thèmes font l'accord des économistes, d'autres en revanche suscitent les controverses.

Les problèmes « d'agence » et l'impossibilité d'envisager toutes les contingences imaginables dans les contrats sont aujourd'hui reconnus comme cruciaux.

La séparation de la propriété et de la direction dans les grandes entreprises est à l'origine d'un problème d'agence (encore appelé problème de « mandat »). En effet, si l'équipe dirigeante a pour mandat d'entreprendre toute action conforme à l'intérêt des actionnaires, elle dispose d'une meilleure information que ces derniers sur l'environnement de l'entreprise : nature des marchés, risque associé aux projets, etc. Cette asymétrie d'information permet aux dirigeants d'orienter en partie la gestion en fonction de leurs propres intérêts, qui peuvent être différents de ceux de leurs actionnaires qui souhaiteraient essentiellement la valorisation du capital investi.

Par ailleurs, la négociation et la signature de contrats complets entre actionnaires et dirigeants entraînent de tels coûts de transaction qu'en pratique seuls des contrats incomplets sont signés. Des dispositifs spécifiques de gouvernement d'entreprise doivent alors être utilisés pour résoudre les problèmes d'agence qui ne peuvent s'appuyer sur la formalisation de contrats détaillés entre actionnaires et dirigeants. Dans cette perspective, on distingue généralement deux modes de contrôle et d'incitation des dirigeants : le contrôle externe, exercé essentiellement par le marché financier ; et le contrôle interne, exercé par les actionnaires et leurs représentants, le conseil d'administration.

La comparaison avec le contexte du secteur public suggère donc à la fois un trait commun, à savoir l'importance des problèmes d'agence à résoudre ; mais aussi deux différences notables : un critère à maximiser de surplus social qui est encore plus difficile à identifier que ne l'est la valeur de l'entreprise, dans le contexte privé ; l'absence des mécanismes de contrôle externe, qui renforcent les enjeux de la mesure de la performance dans le secteur public.

L'examen des débats et controverses concernant les mécanismes de contrôle des dirigeants d'entreprise permet par ailleurs de suggérer la nature des questions à se poser lorsque l'on veut les transposer au secteur public :

- les incitations monétaires : elles coûtent cher à l'entreprise et entraînent une certaine incertitude sur la rémunération des dirigeants de l'entreprise, ce qui n'est pas optimal si ceux-ci ont de l'aversion pour le risque vis-à-vis de leurs revenus. Leur qualité dépend par ailleurs crucialement de la

mesure de performance : doit-on se fonder sur les mesures comptables du profit ou plutôt sur la valeur boursière, sachant qu'il convient de veiller à un certain équilibre dans les incitations, notamment entre incitations de court terme et incitations de long terme ? Par exemple, donner aux dirigeants une participation au profit sous forme de bonus constitue une incitation de court terme. Elle peut conduire les dirigeants à réduire les investissements et à négliger le long terme. Donner aux dirigeants des actions ou des stock-options constitue une incitation de plus long terme, à condition toutefois que les investissements se reflètent bien dans la valeur boursière de l'entreprise.

- le « *monitoring* » : il consiste à intervenir dans l'entreprise pour en accroître la valeur. Il peut être pratiqué par différents acteurs – bien entendu, le conseil d'administration dont c'est le rôle, mais aussi les gros actionnaires, par exemple, des fonds de pension ou du capital-risque ou des gros créiteurs, comme des banques ou des compagnies d'assurance – et revient à créer un surveillant au-dessus des dirigeants. Ce type d'incitations peut soulever différents problèmes. Un premier est celui de collusion éventuelle entre surveillant et dirigeants, qui a été souvent mentionné pour le conseil d'administration, dans différents pays. Un autre type de problèmes soulevés par le « *monitoring* » est celui des sous-incitations à exercer celui-ci, car un noyau dur, par exemple, ne recevra qu'une part du supplément de valeur résultant de son action.

3. L'organisation de l'évaluation

Incitations pour gérer un bien commun à Cambridge en 1574 : les prescriptions du legs Parker. Suivant ce testament, les ouvrages composant le legs devaient être logés à Cambridge sous la garde du doyen du collège « Corpus Christi et de ses successeurs à perpétuité », sous les conditions suivantes :

Le six août de chaque année, les doyens des collèges de Gonville et Caius, et de Trinité doivent réaliser un rapport d'audit indépendant... Si lors de l'inspection annuelle, l'un des ouvrages ou manuscrits manque, Corpus Christi fait l'objet d'une pénalité fixée par les deux superviseurs en exercice, de manière à ce que le fonds soit accru à l'issue de la future inspection annuelle.

En cas d'infraction, Corpus Christi perdra non seulement les assiettes d'argent mises en caution chez le susmentionné Matthew, qui seront transférées ainsi que la garde à Gonville et Caius Collège. En cas de négligence similaire de celui-ci, le transfert serait réalisé au bénéfice de Trinité, et retournerait à Corpus Christi, dans les conditions initiales, si ce dernier est défaillant...

Cité par le Journal of Political Economy.

Ces difficultés du « *monitoring* » trouvent directement un écho lorsque l'on réfléchit sur l'organisation de l'évaluation dans le secteur public.

3.1. Éléments d'analyse

3.1.1. Les enjeux

Compte tenu des contraintes informationnelles recensées ci-dessus, il n'est pas surprenant qu'un élément particulièrement notable des auditions réalisées pour la préparation de ce rapport soit l'accent mis sur les questions de méthodes et d'organisation de l'évaluation. À cet égard, Tony Atkinson a mis en avant le besoin de développer un processus qualité sur les données, ceux qui fournissent les « chiffres » devant avoir pleinement conscience des enjeux associés. Les difficultés méthodologiques pour passer d'indicateurs d'activité à une véritable mesure du service rendu le conduisent à admettre la possibilité de plusieurs systèmes d'indicateurs, dès lors que ceux-ci sont cohérents et que leurs développements respectent des cahiers des charges prédéfinis.

La difficulté – on y reviendra – de ce passage entre « activité » et « résultat » ou service résulte de ce que :

- lorsqu'il s'agit de services à l'utilisateur, comme c'est le cas pour l'éducation ou la santé, il n'y a pas en général de référence marchande directe pour évaluer le service, et que, surtout la dimension qualité est très importante ;
- pour les services plus régaliens (sécurité publique, collecte de l'impôt, contrôle de la réglementation...), le service fourni s'apparente plutôt à un service « d'intermédiation », et donc à évaluer comme tel, en coût d'opportunité.

En tout état de cause, ceci suggère que la mesure de la productivité publique restera toujours un exercice délicat, et que l'on ne peut donc passer mécaniquement de l'exercice de mesure, au diagnostic et aux conséquences qui en sont tirées ensuite. En revanche, on peut, partant de ce que devrait être le contenu du diagnostic, identifier les besoins d'information nécessaires pour l'établir.

De son côté, Bernard Belloc (ancien vice-président de la Conférence des présidents d'université) a mis l'accent sur l'organisation souhaitable de l'évaluation, en insistant sur la nécessité de s'assurer de la compétence des évaluateurs, de leur absence de conflits d'intérêts, et du besoin de « certification » du processus. Il souligne aussi le besoin en amont d'affirmer les objectifs de l'institution. Incidemment, ceci souligne que la démarche de performance et de renforcement de l'éthique dans le service public, par la notification, en particulier, des conflits d'intérêt, ne sont pas des choses contradictoires. Au contraire, la revue de l'OCDE (2000) sur cette question soulignait que la fonction publique est dépositaire de la confiance du public. Les citoyens attendent des fonctionnaires qu'ils servent l'intérêt général en faisant preuve d'impartialité, et en gérant quotidiennement les ressources publiques de manière appropriée. Un service public équitable et fiable inspire confiance au public. Ceci crée aussi un environnement favorable pour les entreprises, ce qui contribue au bon fonctionnement des marchés et à la croissance économique (*cf.* encadré 4).

4. Renforcer l'éthique dans le service public : principales conclusions^(*)

Les citoyens ont confiance dans leurs institutions publiques lorsqu'ils savent que les charges publiques sont utilisées dans l'intérêt général.

Ceci suppose notamment de :

- définir une mission claire pour le service public ;
- préserver les valeurs tout en s'adaptant au changement ;
- permettre aux fonctionnaires et aux citoyens de faire état des infractions ;
- inclure les mesures favorisant l'intégrité à l'ensemble de la gestion et coordonner celles-ci ;
- passer de l'application des règles à la prévention ;
- prévoir les domaines susceptibles de poser problème ;
- tirer parti des nouvelles technologies.

(*) Cf. OCDE, 2000.

Finalement, il convient aussi de se défier des systèmes d'évaluation centralisés aisément capturables. À cet égard, le complément de Jean-Luc Gaffard apparaît comme une illustration de ce risque, à propos de la réforme des études supérieures dite LMD (licence/maîtrise/doctorat). Il considère en effet que la multiplication des échelons et l'opacité du système d'évaluation centralisé ont favorisé les comportements opportunistes jouant sur les défauts d'information, avec comme conséquence une occasion manquée d'aller vers l'excellence, et la consolidation des rentes locales et de l'absence de qualité. Il plaide finalement pour une évaluation mieux ciblée et transparente, au niveau des établissements, et sous la responsabilité d'un conseil composé de personnalités extérieures aux compétences reconnues.

Le complément de Martine Perbet (2006), qui s'intéresse plus généralement aux expériences d'évaluation en France, aboutit à un diagnostic très convergent. Outre le fait que le flou de la commande associé à l'absence d'analyse économique a souvent conduit à des produits d'évaluation très hétérogènes, celle-ci constate en effet :

- que la gouvernance de l'évaluation s'est vue accorder peu d'attention ;
- que le modèle recherché pour l'évaluation *ex post* a souvent été celui d'une instance indépendante ayant le monopole de l'évaluation, avec l'espoir qu'émergerait ainsi un produit de qualité incontestable. Elle suggère que mieux vaudrait tabler sur l'ouverture et la confrontation des évaluations.

Ainsi, l'organisation de l'évaluation de la performance apparaît comme est un sujet en soi.

3.1.2. Les problèmes économiques associés à la supervision

Dans la littérature économique, ce problème est associé à la question du « contrôle des contrôleurs ». Celle-ci se pose dans toute institution, et apparaît centrale pour la compréhension de l'organisation des firmes et des relations entre managers et financiers. De manière générale, la gestion publique apparaît fragile à cet égard, en l'absence d'un mécanisme analogue aux procédures de supervision exercées par le secteur bancaire sur les firmes (et ce dans des conditions où celui-ci est par ailleurs soumis à des contraintes de réputation l'obligeant à exercer efficacement ce rôle de superviseur).

Dès lors, la fonction de supervision est soumise elle-même à des problèmes d'incitation et de risque de capture, dont on peut donner l'intuition revenant au modèle principal-agent de référence de Laffont et Tirole (1986). Celui-ci considère le problème de régulation d'une agence privée par un principal en situation d'information asymétrique. L'asymétrie porte à la fois sur l'observation de la capacité intrinsèque de l'agence par rapport à la tâche qui lui est confiée, et sur celle de l'effort fourni par celle-ci pour la réaliser. En revanche, le modèle suppose que le principal est apte à observer une combinaison de ces deux termes, au travers des coûts comptables de l'agence.

Dans ce cas, la divergence d'objectifs entre le principal et l'agent résulte du fait que ce dernier est une firme privée, bénéficiant de droits exclusifs. Pour notre propos, les enseignements de ce modèle sont cependant intéressants, car ils illustrent certaines conséquences de l'asymétrie informationnelle que l'on retrouve dans toute situation d'agence publique ayant des motivations, préférences ou talents propres qui font que ses objectifs ne sont pas parfaitement identiques à l'intérêt général, et que la gestion publique doit remettre en ligne, avec les problèmes de supervision associés. Il reste vrai, comme on l'a amplement montré, que le modèle à considérer pour la gestion publique devrait être plus complexe compte tenu de la multiplicité des tâches et des principaux.

Si il n'y avait pas d'asymétrie d'information, le principal, connaissant l'efficacité intrinsèque de son agence, pourrait en contrôler l'effort, en fixant le montant de remboursement des coûts qu'il accepterait, selon son type. Quel que soit le niveau d'efficacité intrinsèque de l'agence, le contrat proposé dans ce cadre serait associé à une absence totale de rente octroyée à celle-ci.

Si le principal conserve ce type de contrat, mais qu'il est en fait en situation d'information asymétrique, l'agence du type efficace aura plutôt intérêt à relâcher son effort, pour bénéficier du remboursement prévu pour l'autre type. Celle-ci bénéficiera alors d'une rente « informationnelle », traduisant le fait que le principal ne sait pas au fond trancher, quand il constate que l'agence a des coûts élevés, si ceux-ci résultent de la difficulté de la tâche pour son agence, ou de son manque d'effort. L'asymétrie d'information a alors un coût pour la société, dû au fait que le principal doit arbitrer entre laisser des rentes ou inciter à l'effort, avec comme cas polaires : les con-

trats de type remboursement des coûts (« *cost-plus* ») qui ne laissent pas de rentes mais n'incitent pas l'effort et les contrats à prix forfaitaires (« *price-cap* ») qui incitent à l'effort mais en laissent tout le gain à l'agence concernée.

Poursuivant plus avant, différents enseignements peuvent être tirés de ce modèle pour notre propos. Les premiers ne font que réexpliquer l'idée que les contraintes informationnelles amènent à réviser la conception de l'intervention publique :

- l'application de règles de remboursement des coûts conçues dans un contexte d'information parfaite s'avère complètement inappropriée en termes d'incitations à l'effort si cette hypothèse n'est pas vérifiée, comme cela vient d'être rappelé ;

- on peut cependant optimiser l'arbitrage entre rentes et efficacité, grâce à un « menu de contrats », qui distordra le niveau d'effort dans l'hypothèse d'une agence de type moins efficace, de manière à limiter la rente informationnelle laissée dans le cas où elle est de l'autre type⁽¹⁰⁾. La comparaison entre les deux contrats s'interprète comme un taux de retour sur le gain de productivité.

Mais il faut bien comprendre que tant que cette rente informationnelle demeure, l'obtention de l'information pertinente sera toujours difficile, car ce n'est pas l'intérêt de l'agent de la révéler (tout mécanisme optimal à cet égard conduisant à la révélation de l'information privée au prix d'une rente)

Il en résulte qu'un système de supervision aura alors une grande valeur, en permettant de relâcher ces contraintes informationnelles, et cela est bien l'objectif implicite au développement de systèmes d'indicateurs de performance. Mais sa mise en place rencontrera aussi des problèmes d'incitations, le superviseur et l'agence (si elle est du type le plus efficace) ayant intérêt à un comportement collusif, préservant la rente informationnelle pour se la partager.

Dans ce cas (Laffont, 1999), le principal n'a plus que trois solutions :

- diminuer les enjeux de collusion, c'est-à-dire accepter une gestion publique rigide, ne stimulant pas les compétences et talents de ses membres, mais ne leur laissant pas non plus de rentes ;

- augmenter (ce qui a un coût) les coûts de transaction de la collusion ;

- créer (ce qui a aussi un coût) des incitations pour que les superviseurs rendent compte d'informations qui obligeront l'agent à fournir des efforts plus élevés, sans lui laisser de rente ; en d'autres termes « payer le prix d'une administration qui bénéficie de la discrétion nécessaire à l'adaptation aux circonstances locales, et aux spécificités des administrés ».

(10) Ce modèle fournit une bonne intuition des arbitrages à réaliser des lors que la rente est socialement couteuse bien qu'il repose sur des hypothèses spécifiques sur la fonction de coût.

La première solution correspond à un service public de maigre qualité. On verra ci-dessous que ce type de régulation correspond peu à la réalité, en termes de discrétion laissée aux agents. Inutile de rappeler par ailleurs que les coûts de transaction envisagés dans la seconde solution doivent évidemment être en rapport avec les modalités de la collusion, qui le plus souvent se joue dans l'influence et la durée, et non dans un partage immédiat de cette rente.

Seule la réintroduction d'une forme de concurrence pourrait donc permettre d'échapper à ces dilemmes entre rentes et incitations. Mais les raisons qui justifient l'intervention publique, par exemple l'existence d'économies d'échelle, font que celle-ci ne peut être, en général, une concurrence sur le marché. Toutefois, d'autres formes de concurrence sont envisageables : *concurrence ex ante* « pour le marché », suivant les idées originelles de Chadwick au XIX^e siècle (Mougeot et Naegelen, 2005) ou *concurrence par comparaison*, le niveau de rémunération forfaitaire accordé à l'agence, étant calibré, sur la base des coûts observés dans des agences exerçant dans des conditions similaires, sur d'autres territoires.

Ces contraintes expliquent le regain d'intérêt pour les partenariats public-privé, qui constituent en effet un moyen d'externaliser la réalisation du service public, dans des conditions contractuelles mettant l'accent sur la qualité du service fourni, et avec un engagement de durée suffisante pour arbitrer efficacement entre coûts d'investissement et coûts de fonctionnement, ou pour gérer efficacement les risques d'interfaçage entre les différentes composantes d'un projet. Dans ce cas, la question des incitations données au superviseur se déplace vers celle des incitations données à celui qui sélectionne le prestataire, problème qui dépasse de beaucoup le formalisme des règles de sélection.

S'agissant de la *concurrence par comparaison*, celle-ci peut s'exercer à différents niveaux : d'abord entre unités de la même organisation, ou entre différents services publics exerçant des activités similaires, État et collectivités locales, par exemple, lorsque c'est le cas, ou entre services publics et services privés, comme on l'a vu pour certains services publics locaux ou entre hôpitaux et cliniques. La tarification à l'activité mise en œuvre actuellement dans le système hospitalier en est un exemple. D'autres secteurs pourraient adopter des mécanismes analogues notamment le secteur de l'enseignement primaire, secondaire et supérieur à condition de mettre en place des indicateurs de performances fiables et acceptés. En l'absence de possibilité de comparaison nationale, les références étrangères sont évidemment irremplaçables.

Plus généralement, on a toujours intérêt à élargir le champ de la comparaison, mais ce faisant, le besoin de standardisation des données pour éliminer les écarts dus à des facteurs d'environnement augmente aussi et il convient donc de s'y atteler avec rigueur.

3.2. Enseignements et propositions

3.2.1. Principes généraux

Ces principes peuvent être résumés par sept propositions :

- le travail à réaliser pour construire un système d'indicateurs pertinent est certes conséquent, mais les difficultés méthodologiques à résoudre ne sauraient constituer un obstacle à l'objectif d'évaluer la performance de l'action publique, à partir d'une approche fondée sur la mesure des services fournis ;
- une attention particulière doit être portée à l'évaluation de la qualité, à la fois parce que c'est une dimension essentielle des services publics, et parce que le débat public sur ce thème est un moyen puissant de stimuler la performance ;
- les indicateurs correspondants doivent systématiquement être rendus publics ;
- les évaluations doivent faire l'objet de suite, incitant à la performance, ce qui implique de laisser un « retour » aux plus performants. L'approche « compassionnelle » est inutile. Mieux vaut chercher plutôt à définir les mesures d'accompagnement en termes de formation, d'encadrement ou de tutorat, rendant accessible à tous un processus de progrès ;
- la production des données doit faire l'objet d'un processus qualité allant jusqu'à leur certification ;
- l'évaluation n'est pas un exercice mécanique, et peut-être corrompu comme le montrent les expériences hospitalières et universitaires. Il convient donc de l'organiser en conséquence, en évitant les collusions qui confortent les rigidités existantes ;
- l'approche en termes d'évaluation par comparaison (« *benchmarking* ») est un moyen très utile dans cette perspective qui doit être systématiquement développée.

Les indicateurs de performance élaborés récemment par l'Institut de la gestion déléguée (cf. encadré 5) dans le domaine de l'eau sont proches de ce cahier des charges, et des indicateurs similaires y sont en cours de développement pour les transports publics et le service public des déchets.

On dispose par ailleurs de références internationales pour l'éducation, la santé, la recherche, voire la sécurité publique.

L'exemple de la santé est utile pour préciser ce qu'apportent les différents types d'indicateurs au processus d'évaluation. Typiquement, on dispose, en ce domaine, d'abord d'indicateurs généraux sur la situation sanitaire et les coûts de santé associés : espérance de vie corrigée de l'incapacité, inégalité de la distribution de cette variable au sein de la population, accessibilité au système de soins et discriminations éventuelles au regard de celle-ci, poids des dépenses de santé et inégalité éventuelle de son financement. Ces données, que publie l'OMS, sont cependant des données « d'état »,

5. Régie, marché, partenariat, délégation : quelle compétition pour l'amélioration du service public ? Comparabilité, transparence, réversibilité^(*)

La question du mode de gestion entre régie, affermage et concession se pose avec une acuité particulière en France pour les services publics de l'eau, des déchets, des transports urbains...

Constatant les difficultés techniques pour comparer entre eux les modes de gestion et soucieux de garantir leur libre choix, l'IGD a mis en place en 2004 un groupe de travail sur les conditions de compétition entre modes de gestion du service public : la régie directe, le marché, le contrat de partenariat et la délégation. Le groupe de travail présidé par Christian Babusiaux, Conseiller maître à la Cour des comptes, a cherché à déterminer les conséquences économiques de ces différences. Il a ainsi examiné les conditions dans lesquelles pourrait s'exercer une compétition entre les divers modes de gestion, il en a recensé les difficultés, les a hiérarchisées et enfin a dégagé des propositions. Celles-ci ont pour objectif de permettre l'exercice d'un choix non pas dans les conditions uniformes mais exempt de différenciations non identifiées ou injustifiées. Les propositions dégagées par le groupe de travail dans son rapport final visent ainsi à créer les bases d'une comparabilité, pour que la décision publique soit au mieux éclairée.

L'objectif est la constitution pour chaque secteur d'une grille d'analyse simple, objective, et efficace pour la comparaison des situations dans le temps ou entre collectivités organisatrices. Les indicateurs doivent être en nombre limité, accessibles et fiables dans la durée.

Le groupe de travail « Eau et assainissement » qu'a installé l'IGD en 2003, s'est évertué sous l'égide de Jean-Louis Coppeaux à identifier dans ce secteur douze indicateurs de performance représentatifs que les professionnels se sont engagés à déployer sans attendre sur le plan national. Une démarche identique est suivie dans le domaine des transports (avec le groupe de travail présidé par Louis Nègre), des déchets, de l'énergie, de la restauration collective..., de nature à installer la culture de l'évaluation dans le choix du mode de gestion des services publics locaux.

Les travaux de ces groupes de travail ont pris une dimension nouvelle avec la parution le 14 mars dernier d'un décret sur le contenu du rapport du délégataire (article R.1411-7 CGCT). Ces rapports devront en effet, à partir de janvier 2006, comporter une analyse de la qualité du service, qui est « notamment appréciée à partir d'indicateurs proposés par le délégataire ou demandés par le délégant et définis par voie contractuelle ».

(*) IGD, 2005.

très globales, qui ne permettent que de fournir des « indications » sur les performances relatives des pays. D'une part, les pondérations qui sont utilisées pour établir ces comparaisons sont discutables. D'autre part, la performance ainsi mesurée dépend d'un ensemble de facteurs (en particulier des

comportements vis-à-vis du risque), au sein desquels la gestion de systèmes de santé n'est donc qu'un élément.

Des méthodes ont donc été développées, plus avant, d'une part pour identifier les contributions des différents facteurs et estimer l'impact des actions possibles, d'autre part pour estimer combien la société est prête à dépenser pour améliorer ces indicateurs de santé. Ces éléments permettent d'évaluer les politiques en termes de bilans coûts-avantages, par exemple ici les règles souhaitables de prise en charge des différents traitements.

Celles-ci déterminent évidemment l'efficacité sociale du système de soins qui les met en œuvre. Mais cette efficacité dépend aussi des conditions dans lesquelles ceux-ci sont fournis, ce qui conduit à rechercher d'autres indicateurs pour caractériser l'offre de soins : durée de séjour et coûts par pathologie, qualité des soins, risque de maladies nosocomiales, etc. Là encore, des instruments de mesure peuvent être élaborés, étant entendu que, pour être pertinents, ils doivent rendre compte du recrutement de l'établissement considéré en termes de gravité des pathologies traitées (et donc de l'éventualité d'une sélection des risques). Par ailleurs, s'agissant d'activités où il y a nécessairement une grande variabilité des cas traités, la mesure de la performance ne peut se faire qu'en moyenne, sur une période d'activité. Sous ces conditions, des différences entre établissements concernant, par exemple, le pourcentage de patients traités en moins de x jours pour un type de pathologie standardisé, reflètent bien les écarts de performances entre ceux-ci. On peut ensuite bâtir sur ces bases une tarification des actes. Ceci nécessite cependant de s'interroger sur les incitations qui seront ainsi fournies. Même si l'indicateur suggéré ci-dessus reflétait bien la performance, il ne pourrait encore être utilisé directement à cette fin, car il générerait un effet de seuil inacceptable (à x jours). Ce qu'il faudrait rechercher serait donc un barème dégressif, mais plus continu, de la rémunération de la journée supplémentaire.

En l'absence d'analyse *a priori* explicite des modifications de comportements qui seront générées par l'introduction de nouvelles formes de rémunération à la performance, le risque est très grand que celles-ci conduisent dans un premier temps à beaucoup de reports non maîtrisés, comme on a pu les observer dans la gestion hospitalière (transferts de malades, choix d'hospitalisation ou de prestataires pour les examens incompréhensibles...), qui discréditent les réformes avant même qu'elles ne puissent commencer à porter leurs fruits.

La recommandation en faveur de ce type d'analyse est valable de manière générale en matière d'évaluation des réglementations, où l'on sait que des phénomènes qualifiés de dérives ou de contournements ne font que refléter l'adaptation rationnelle des comportements à un nouvel environnement. Mais elles revêtent une acuité particulière dans la gestion publique, car les systèmes de rémunération y demeurent frustes (ou de « nième rang »), si bien que beaucoup de variables d'adaptation peuvent avoir été mal anticipées. À titre d'exemple, on peut être étonné du raffinement que l'on met sur

le formalisme des marchés publics, sans se poser la question de savoir si l'acheteur concerné a véritablement une incitation à minimiser les coûts. Autre exemple, en matière de gestion des effectifs, une politique se focalisant sur le taux de vacance conduit souvent à générer des comportements de rétention qui vont à l'encontre des objectifs ultimes visés, de réduction ou de redéploiement.

Bien évidemment, la description qui précède n'a qu'une valeur heuristique, pour identifier l'enchaînement des différents types d'indicateurs à considérer. Elle permet de suggérer que cette démarche est peu ou prou transposable dans les autres secteurs de l'action publique.

Notons que l'on s'est principalement situé ici au niveau de la performance des services ou établissements, l'idée étant comme cela l'a été indiqué, d'une part que la performance publique est en général une performance « d'équipe », qui doit être nettement distinguée de l'évaluation individuelle, d'autre part que les incitations données aux agents doivent tenir compte des leviers d'action dont ils disposent effectivement.

Ceci n'empêche pas que ce volet est aussi important, car il n'y a pas de bon orchestre si ses instrumentistes ne maîtrisent par leur partition. Des outils existent déjà en ce domaine avec les entretiens annuels d'évaluation, la notation, et ses conséquences sur l'avancement.

Le premier souci à avoir est donc sans doute de les rendre plus effectifs, grâce, par exemple, à un meilleur suivi infra-annuel des agents. Une autre priorité est par ailleurs d'introduire les différenciations de rémunération qui sont nécessaires pour attirer les agents dans certaines régions ou sur les tâches où le déséquilibre entre les besoins et les offres est avéré. Ensuite, la question d'une meilleure prise en compte du mérite dans la rémunération se pose sans doute avec le plus d'acuité pour les personnels qui sont les moins touchés par les incitations à l'avancement. À cet égard, la faiblesse de différenciation affectant les cadres B et C de l'administration peut être jugée bien inéquitable, au détriment de ceux qui donnent le plus au service public. Au-delà, il convient de rappeler la nécessité, lorsque l'on renforce les incitations dans la rémunération individuelle sur un critère particulier, de bien s'assurer que les tâches collectives ne seront pas délaissées.

3.2.2. Propositions

S'agissant du risque de « corruption » du processus de réforme, le secteur de la santé est, comme on l'a vu, marqué par la récurrence de réformes dans lesquelles beaucoup d'espoirs avaient été placés, qui ont ensuite été déçus, notamment parce qu'elles avaient ignoré (Mougeot, 1999 et Bureau, 2002) la dimension microéconomique.

L'expérience doit être retenue pour la réforme des lois de Finances, (LOLF) et les stratégies ministérielles de réforme (SMR). Certes, celles-ci ont permis d'engager un processus de réforme structurelle de la gestion

publique, avec comme progrès immédiats : une redéfinition et une clarification des missions des services, associées à la définition des programmes et des actions de la LOLF, l'introduction d'une culture de performance, avec le travail sur les indicateurs, la modernisation des dispositifs d'évaluation des agents, avec la généralisation de véritables entretiens annuels. Ce processus a aussi été l'occasion de révéler certains modes d'organisation qui ne sont plus tenables, compte tenu de l'évolution de la décentralisation ou de la libéralisation de certains marchés, par exemple.

Mais le développement d'indicateurs globaux, au niveau des programmes de la LOLF ou au niveau macroéconomique de la comptabilité nationale, ne devrait être considéré que comme la partie la plus émergée d'un processus d'évaluation, qui se joue d'abord au niveau microéconomique des agences ou des établissements. De même que l'on sait depuis Adam Smith que l'efficacité économique dépend des conditions dans lesquelles chaque individu, en poursuivant son propre intérêt, travaille à celui de la société, sachant que « l'individu ne se fixe en général pas pour but de promouvoir l'intérêt public, pas plus qu'il ne sait dans quelle mesure il le favorise... », la performance publique est aussi la résultante de celle de ses unités de gestion élémentaires, et des conditions dans lesquelles cette convergence d'intérêt est réalisée au sein de celles-ci.

De manière plus précise, on observe dans la plupart des services publics une très grande hétérogénéité des performances, avec souvent une queue de distribution importante d'unités à faible rendement apparent (hôpitaux ou maternités sous utilisées, classes à très faible effectif, lignes ferroviaires à faible trafic, petits bureaux de poste, etc.). Mais leur cumul n'explique en général qu'une partie des résultats globaux, si bien qu'il faut véritablement s'intéresser à l'ensemble des distributions, pour comprendre l'hétérogénéité qui peut être justifiée par des conditions d'environnement et celle qui traduit les écarts à la frontière de productivité qu'il faudrait résorber.

Un autre argument en faveur d'une approche microéconomique réside dans le caractère local des services publics, et vis-à-vis desquels se développe un véritable consumérisme, y compris pour les services les plus régaliens comme la sécurité. Enfin, et c'est l'évidence, de même que les risques doivent être alloués aux niveaux les plus à même de les gérer, la productivité devrait d'abord être mesurée aux niveaux de gestion les plus à mêmes d'y contribuer.

La proposition qui suit découle de cette analyse.

Il convient de compléter le travail engagé avec la LOLF sur les indicateurs au niveau de l'information du Parlement, en développant la mesure de la performance au niveau microéconomique des services, agences ou établissements publics. C'est en effet à ce niveau que celle-ci se joue, et il n'y a pas de performance sans une véritable évaluation (« *accountability* ») de l'action publique.

Celle-ci devrait impérativement renseigner, au-delà de l'information sur les moyens et l'activité, sur les résultats et services fournis au public, en permettant la comparaison avec des établissements opérant dans des conditions comparables (ou rendues telles).

Ces résultats devraient être établis de manière indépendante, certifiés, et rendus publics.

Sur la base de ceux-ci, des propositions d'actions (« suites ») devraient systématiquement être présentées à leurs instances d'orientation, avec un dispositif de suivi.

Chacun des éléments de l'énoncé ci-dessus pourrait ensuite être décliné (*cf.* annexe) :

- en suggérant de favoriser l'innovation sur chacun des points visés ;
- en rappelant les objectifs (absence de conflits d'intérêt et implication des évaluateurs, par exemple) et le besoin de gérer les transitions ;
- en recommandant de développer la recherche en économie de la gestion publique, les travaux tels que ceux d'Alain Trannoy et Robert Gary-Bobo sur l'éducation ou certaines recherches sur l'organisation sanitaire apparaissant comme une exception trop rare.

Cependant, il nous semble qu'il faut d'abord insister sur le fait qu'un tel processus d'évaluation ne peut prendre son essor que si la cible visée est bien affirmée, et que les conditions d'effectivité sont en place. La comparaison avec la situation américaine, en matière d'études d'impact des réglementations, notamment environnementales, ou d'évaluation contradictoire de la gestion publique, avec l'implication à la fois du Congrès, du contrôle (*General Accounting Office*), et de l'exécutif (avec l'*Office of Management and Budget*) peut ici servir de référence.

Surtout, il faut vraiment insister sur la *nécessité de se prémunir contre les risques de capture du contrôle*, ce qui conduit à écarter très nettement le modèle d'instance centralisée, ayant le monopole de l'évaluation. À titre d'exemple, Proper et Wilson (*op. cit.*) suggèrent le développement d'objectifs basés sur des sources d'information indépendantes, alternatives, telles que par exemple, l'enquête sur les crimes britanniques pour fixer des objectifs aux autorités de police, ou l'utilisation d'enquêtes générales sur les ménages pour évaluer la santé des personnes habitant dans une zone. À l'appui de cette recommandation, ils mentionnent notamment le besoin d'indicateurs non « corruptibles », non sujets à la manipulation par les individus dont les actions sont mesurées. Ils insistent ainsi sur la divulgation publique des informations de performance, en notant, accessoirement, que l'utilisation de sources d'information diversifiées pousse les organisations à se concentrer sur ce qui est réellement important (par exemple, la prévention des crimes ou la prévention des maladies), ce qui devrait également les encourager à se demander ce qui est réellement important.

Mais le problème est de portée générale, et se pose, par exemple, en matière d'incitations individuelles. Le fait que l'on plaide pour une introduction raisonnée d'éléments incitatifs variables de rémunération (parce qu'il faut tenir compte de la motivation des agents publics, et surtout parce que des incitations trop « puissantes » risquent d'avoir les effets pervers détaillés ci-dessus) n'empêche pas que l'incitation individuelle est extrêmement importante. Le fait, par exemple, que des primes qui devraient rémunérer une activité particulière puissent être distribuées de manière uniforme, conduit toujours à des impacts très visibles (et très contagieux), que l'on pourrait corriger simplement en théorie en revenant à l'application des règles. Mais encore faut-il que les « superviseurs », c'est-à-dire ici les échelons hiérarchiques immédiatement supérieurs soient incités à exercer ce rôle, et non préférer l'arrangement plus facile consistant à admettre (par exemple), que compte tenu des difficultés que rencontre l'agent, il serait injuste de lui supprimer les heures supplémentaires bien qu'il ne les fasse pas.

Comme on le voit, le problème de la mesure de la performance ne peut ainsi être distingué de ceux que rencontre la réforme de l'État en général.

4. Évaluation des performances et réforme de l'État

Le développement étatique planifié a échoué. Mais aussi le développement sans État (...).

Construire un État effectif pour favoriser un développement durable et la réduction de la pauvreté ne sera pas facile. À chaque fois, il y aura des intérêts établis pour garder l'État comme il est, quel qu'en soit le coût social pour le pays dans son ensemble (...).

Introduction au vingtième rapport de la Banque mondiale (1997)

Le sujet de la réforme de l'État dépasse évidemment le champ de ce rapport. Mais on voit bien notamment dans la contribution de Jean-Luc Gaffard (2007), que la question des frontières des agences, du degré de discrétion laissé à celles-ci ou aux agents, de l'articulation des agendas du politique et de la gestion ou des relations avec les responsables budgétaires sont déterminantes. En d'autres termes, la mesure de la performance au niveau microéconomique ne peut porter ses fruits que si elle s'exerce dans un environnement général favorable, notamment du point de vue incitatif. Ceci suppose au moins *trois conditions*, dont l'analyse est esquissée ci-dessous.

4.1. La possibilité de gérer : organisation des agences et discrétion laissée aux agents

Les orientations qui précèdent amènent en effet à s'interroger sur les entités les plus pertinentes pour développer cette « *accountability* ». À l'étranger, notamment dans les pays anglo-saxons, ceci correspond à la no-

tion « d'agence » ou « d'office », entités ayant beaucoup plus de responsabilités de gestion que nos « services » mais beaucoup moins « d'indépendance » que nos établissements publics (administratifs), qui souvent apparaissent piètrement « contrôlés et équilibrés », et entretiennent souvent des relations conflictuelles avec leurs tutelles. Une réflexion devrait donc être engagée pour refonder les règles en ce domaine, en intégrant les définitions des missions données ci-dessus et les contraintes de mesure de la performance. En particulier, celles-ci doivent leur fixer des missions aussi définies que possible, pour que les agents sachent ce qu'ils ont à faire et que leur performance soit mesurable. Simultanément les besoins de coopération entre agences doivent être identifiés et organisés. Des règles sont nécessaires aussi pour établir les bases de leur bonne gouvernance, et éviter ainsi que les statuts de ces entités soient affectés par des éléments contingents à leur création, voire corrompus à ce moment par la nécessité de laisser des gages à ceux qu'elle obligerait à se réformer. Une telle réflexion est d'autant plus urgente que les catégories intermédiaires (Groupements d'intérêt public ou scientifique, services à compétence nationale, etc.) n'ont pas pris leur essor et ont plutôt constitué des solutions à la marge qu'une nouvelle référence.

S'agissant de la discrétion laissée aux agents, on laisse curieusement *beaucoup de marge de manœuvre aux gestionnaires publics dans le choix de leurs interventions ou des services offerts, et on les soumet à beaucoup de contraintes dans la gestion des moyens*. On est plus éloigné d'une situation où l'offreur public choisit à la place de son client. Or, s'agissant de services, où l'enjeu de qualité est fort, l'offre publique devrait au contraire faire l'objet d'une démarche systématique de certification, pour répondre aux attentes réelles de l'utilisateur, et ce au moindre coût.

Sur les usagers, la situation est inverse, alors que l'autonomie de leur gestion est fondamentale. Il serait par exemple justifié d'ouvrir le débat sur l'autonomie de certains établissements, notamment dans le domaine de l'enseignement supérieur et de la recherche. À tout le moins, des marges de gestion appropriées doivent être données.

La constitution des budgets opérationnels de programmes (BOP) et des unités opérationnelles (UO) de la LOLF et leurs protocoles de gestion afférents devraient systématiquement être examinées par rapport à ce critère (ce que suggèrent aussi Alain Lambert et Didier Migaud (2006) pour qui « la direction du budget doit se livrer à un « contrôle qualité » des Projets annuels de performance »). Les modes d'évaluation des corps d'inspection et de contrôle devraient aussi en tenir compte, sachant que ceux-ci ne peuvent se contenter de regarder si les allocations budgétaires ne surcompensent pas les coûts engagés. On a vu en effet que ces coûts sont endogènes si bien que vérifier qu'aucune rente n'a été laissée, c'est s'assurer indirectement que le service n'avait aucune incitation à l'efficacité (sauf si l'on est en information parfaite).

Ceci conduit aussi à s'interroger sur certaines tendances en matière de réorganisation publique. La production publique se situe essentiellement dans le champ des services. Les activités les plus comparables du privé sont

réalisées par des entreprises dont la taille et la diversité des activités sont extrêmement variables. À cet égard, on ne peut qu'être surpris par l'idée que la réforme de l'État consiste à constituer de grosses entités, avec pratiquement une affectation stricte entre un programme au sens de la LOLF, c'est-à-dire un type de service fourni, et une organisation unique. Bien évidemment, il faut réduire la dispersion des services publics, et l'argument de la multiplicité de leurs tâches n'est souvent qu'un argument mis en avant pour échapper à l'évaluation des performances. Pour autant les économies d'échelle mises en avant pour certains regroupements mériteraient d'être bien documentés. L'expérience du secteur privé montre en effet que le bilan des fusions-acquisitions est souvent incertain, notamment lorsque les entités rassemblées ont des cultures trop différentes, ou n'ont pas envie de travailler ensemble. Souvent le véritable argument sous jacent à ce « *Big is beautiful* » est qu'on espère contourner ainsi les obstacles aux redéploiements des effectifs, l'idée étant que, si les redéploiements interstructures sont difficiles, on peut plus aisément réaliser des redéploiements internes. Cette idée ne peut être érigée en principe général d'action, car les rigidités intrastructures peuvent être très fortes aussi. De plus, ce regroupement peut, dans certains cas, accroître le pouvoir de monopole ou de blocage d'une corporation particulière.

L'argument des économies d'échelle est souvent utilisé, car il constitue la meilleure justification pour échapper à toute forme de concurrence. Le souci louable de simplification peut ainsi empêcher parfois l'émergence de structures et de modes d'organisation d'avenir de l'État. Il faut apprécier, au cas par cas, les possibilités de redéploiement progressif au sein d'une même entité ou, au contraire la nécessité de créer, en dehors, les structures dédiées aux missions nouvelles et donc de gérer explicitement le déclin d'entités ayant perdu leur véritable raison d'être, pour une raison ou une autre. La première solution suppose que les compétences dont on envisage le redéploiement soient effectivement appropriées (éventuellement après formation) aux missions nouvelles que l'on souhaite leur voir tenir, que ces missions nouvelles soient économiquement justifiées, c'est-à-dire que l'intervention publique que l'on développe permette la correction d'une imperfection de marché et enfin que l'entité sur laquelle on s'appuie dispose d'un portefeuille suffisant d'activités pérennes. Sinon il vaut mieux combiner la création de structures nouvelles, qu'il faut protéger tant qu'elles n'ont pas atteint la masse critique permettant d'en faire un levier des redéploiements. Par ailleurs, il vaut mieux souvent accepter tôt les surcoûts liés au reclassement des agents affectés par les restructurations plutôt que d'escompter des redéploiements dont l'horizon effectif est en fait très éloigné, voire virtuel.

L'appréciation de la possibilité de gérer met en cause, par ailleurs, les règles de gestion du personnel, qui constitue en effet le principal facteur de production des services publics. Au niveau de la LOLF, les coûts correspondants devraient donc être imputés aux programmes les utilisant effectivement, dans une logique de coût complet ou de « qui décide paye ». L'idée qu'en les concentrant sur les programmes support des ministères on élargit le champ des redéploiements possibles, bute encore sur ce problème d'in-

citations. Une telle situation conduirait en effet à ce que les services utilisateurs n'aient pas de retour sur les gains d'efficacité qu'ils réalisent. Leur intérêt est alors de les cacher... et de réclamer des effectifs au moment où l'actualité rend particulièrement sensible telle ou telle mission vis-à-vis du public.

Le « grand » redéploiement escompté s'avère alors tout à fait virtuel. Il en va de même des réductions d'effectifs gérées en taux de remplacement des départs à la retraite. Si un tel ratio est en effet utile en termes de cadrage macroéconomique, ce ne peut être un instrument de gestion. Au niveau microéconomique, le renouvellement des effectifs d'un service a en effet d'autres causes, qui peuvent de plus être stimulées et permettent ausside gérer activement des formules de rémunération partiellement forfaitisées, dans les conditions discutées ci-dessus. L'affichage du taux de remplacement conduit au contraire à exonérer *a priori* d'effort les services qui ont des effectifs jeunes, même s'ils ont des gisements de productivité, et inciter les autres à chercher les bonnes justifications pour obtenir malgré tout le remplacement de leurs départs. Présenter en coût complet les programmes qui fournissent les services finaux à l'utilisateur n'empêche pas d'ailleurs de faire apparaître aussi – mais autrement – les effectifs sur les programmes support, pour rendre compte de la fonction de gestion des ressources humaines qu'ils exercent. À cet égard, on pourrait imaginer une présentation en partie double de ceux-ci, faisant apparaître en recettes les services fournis aux autres programmes, ce qui obligerait à un dialogue sur les « prix de cession internes » et la répartition des coûts communs.

4.2. Une perspective pluriannuelle contractualisée

Les dilemmes les plus aigus entre rentes et incitations sont de nature dynamique et liés à la capacité d'engagement intrinsèquement limitée de l'État. C'est ce que l'on appelle l'effet de cliquet. L'État voudrait annoncer qu'il rémunérera fortement les efforts de gestion. Mais s'il ne peut s'engager de manière crédible à ne pas utiliser exagérément l'information sur les gisements de productivité qu'il obtiendra ainsi, l'agent (ou le service) anticipe que les objectifs qui lui seront fixés ultérieurement seront encore plus exigeants, auquel cas son intérêt est de ne surtout pas s'engager dans cette voie. L'exemple des budgets globaux hospitaliers (et des politiques d'enveloppe globale) est, à cet égard, particulièrement révélateur de telles stratégies.

Ces idées trouvent une application directe lorsque l'on réfléchit, par exemple, sur les moyens de stimuler les recherches en collaboration avec le secteur privé à l'université ou dans les établissements de recherche. En l'absence de taux de retour suffisant pour les équipes directement contributrices, ces collaborations n'ont aucune chance de prendre leur essor. Mais il faut même pour cela que ce taux de retour soit établi dans la durée, et qu'en aucun cas les budgets alloués à ces équipes soient réduits à proportion, sinon l'incitation disparaît.

Au niveau macroéconomique, ces mêmes mécanismes aboutissent à ce que l'on appelle la contrainte de budget lâche (« *Soft budget constraint* »), résultant du manque de crédibilité de l'État à tenir ses engagements en

matière de déficit ou de dépenses budgétaires. Dans ce contexte, tout ce qui accrédite le défaut de capacité d'engagement de l'État, celui-ci essayant de s'approprier immédiatement les marges révélées par les gestionnaires, est dévastateur, car ceux-ci vont alors créer tous les obstacles pour ne pas s'y « faire reprendre ». De plus, il n'est pas bon pour l'exemple que ceux qui jouent le jeu se trouvent *in fine* perdants. L'utilisation excessive de mesures d'exécution budgétaire « à un coup » peut donc être source de rigidités très fortes.

La crainte que la contractualisation pluriannuelle soit source de rigidités est souvent surestimée, étant entendu qu'il n'y a lieu de contractualiser que lorsque toutes les parties y trouvent avantage, les unes parce qu'ils auront ainsi la possibilité de gérer des projets ambitieux, les autres parce que l'on peut en escompter des gains de performances, le retour étant équitablement partagé. Il est par ailleurs toujours possible d'inscrire des mesures d'ajustement à la conjoncture macro-budgétaire dans les contrats.

Rétrospectivement les désagréments de certains contrats entre l'État et les entreprises publiques correspondent à des cas où soit une partie avait accepté de signer un contrat dont elle aurait pu ou du *a priori* juger trop minces les contreparties offertes, soit le respect du contrat s'est poursuivi alors que l'autre partie avait pourtant rompu clairement ses engagements. En résumé, *l'absence de vision pluriannuelle de la LOLF est une lacune*, qu'il faudra combler. Par ailleurs, pour atteindre ses objectifs, le Budget doit d'abord se demander comment donner des gages de la crédibilité de ses engagements plutôt que de faucher le « blé en herbe ».

4.3. Un marché interne du travail plus flexible

La question de la gestion des ressources humaines publiques est évidemment reconnue comme cruciale. Cependant, là encore l'accent est souvent mis seulement sur les données macro-budgétaires : poids des rémunérations dans les dépenses publiques, renouvellement démographique, impact de l'ARTT, etc.

Le rapport de l'Institut de l'entreprise (Kruger, 2002) est ici intéressant car il va au-delà, en soulignant ce qu'il appelle, le défi de la « banalisation » de la gestion des ressources humaines dans un contexte où de « nouvelles formes de concurrence rendent désormais comparables, dans des secteurs d'activité de plus en plus nombreux, des services publics entre eux ou vis-à-vis des entreprises ».

En tout état de cause, les gains escomptés des réformes de structure ne peuvent se concrétiser sans la constitution d'un marché du travail interne à l'administration plus flexible. Les obstacles à cela sont évidemment énormes, car aucun groupe ne souhaite être soumis à plus de concurrence. Aujourd'hui, *ce marché est cependant très rigide*, se limitant essentiellement à un processus de remplacement des postes vacants. L'idée qui pour-

rait guider la réflexion est qu'il faut bien mieux identifier les deux côtés de l'offre et de la demande de travail de ce marché, et créer les instruments pour une meilleure efficacité de l'équilibre. Trois types de mesures, par exemple, pourraient *favoriser cette fluidité* :

- *fiches de poste*. De même que la gestion budgétaire tend à privilégier l'approche en base zéro, par rapport à celle des services votés, l'idée que toute publication devrait être associée à une redéfinition du poste devrait prévaloir. Sinon c'est toujours l'organisation passée qui sert de référence, et se reproduit ;

- *obligation de mobilité*. Le fait de rester plus de cinq ans, par exemple, sur un même poste devrait relever de l'exception. Certes on objectera qu'un enseignant ou un spécialiste ne vont pas changer de discipline. Mais le premier peut changer de niveau de classe, et les conditions d'exercice de sa discipline peuvent avoir évolué justifiant de reconfigurer le poste du second. L'expérience montre que ces arguments sont souvent un alibi pour ne pas aborder de front les problèmes de gestion de personnel ;

- *ouverture systématique des fiches de poste à l'ensemble des corps susceptibles de fournir les compétences appropriées*. De cette manière on décroïsonne les sous-marchés et on stimule une concurrence, ce qui permettrait aux différents corps de mieux orienter leurs formations ou leur gestion des carrières. La circulaire du 2 janvier 2006, qui envisage, à titre expérimental, la mise en place de bourses de l'emploi territorial, s'inscrit dans cette perspective. Mais le sujet est d'importance générale et ne peut donc être résolu seulement par des ajustements marginaux. Un certain nombre d'instruments techniques sont donc à mettre en place pour faire fonctionner un tel dispositif pour respecter notamment les principes d'égalité. Mais les solutions existent (transformation d'emplois, articulation entre première liste au champ plus ouvert et seconde liste n'opérant que sur les sous-marchés, mise en place d'un suivi plus personnalisé pour les agents ayant du mal à trouver une bonne affectation, ce qui serait un « plus » pour les personnels, par rapport à la situation actuelle, etc.).

5. Conclusion : des services et agences rendant compte des services rendus dans des conditions permettant la comparaison des performances

Les problèmes *d'information asymétrique* entre les différents acteurs, et les coûts de révélation de l'information en résultant, sont aujourd'hui reconnus comme le cœur de l'économie des organisations, avec comme dilemme incontournable, en l'absence de mécanismes concurrentiels, des arbitrages entre rentes et incitations ou risques et incitations.

Des *instruments de rémunération* (des agents ou des services) bien conçus permettent d'atténuer ces difficultés. Les instruments optimaux mêlent des éléments de rémunération forfaitaire et des éléments variables dépen-

dant des performances. Mais ils ne peuvent être trouvés sans garder à l'esprit que l'on ne peut tout avoir : permettre aux managers d'avoir le droit de gérer est nécessaire si l'on veut qu'ils utilisent au mieux leurs compétences au service de la performance, et en particulier dans la gestion des aléas ; la contrepartie, nécessaire, est de laisser aux services un *retour suffisant sur les gains d'efficacité* qu'ils obtiennent. Ceci ne signifie pas que l'on est prêt à laisser n'importe quelle rente, sachant que celles-ci sont coûteuses et favorisent les collusions. Mais il faut une *réflexion sur les rentes* que l'on juge utiles pour accroître les incitations, et tirer pleinement les conséquences des cas où l'on a préféré choisir finalement des systèmes peu incitatifs.

L'expérience des modifications de rémunération des opérateurs fournissant des services publics, dans le cadre de processus de libéralisation de certains marchés, a démontré la pertinence de ces réflexions théoriques. Généralement, les formules traditionnelles de remboursement des coûts, qui ne laissent pas de rentes mais ne comportent pas non plus d'incitations à la maîtrise des coûts, ont été remplacées à cette occasion par des formules de rémunération forfaitaire, qui ont les caractéristiques inverses. Les effets ont souvent été tout à fait spectaculaires sur la productivité opérationnelle. Il convient maintenant de les appliquer aussi à notre gestion publique.

Dans ce contexte, la mesure de la performance revêt une importance capitale. Le travail engagé en ce domaine avec la mise en place de la LOLF est donc essentiel. Mais il doit s'inscrire dans une perspective plus large de construction d'un dispositif complet de « *reporting* » ou d'« *accountability* », dont l'unité de base est microéconomique. Ceci constitue, avec la nécessité de se défier des systèmes centralisés d'évaluation, aisément capturables, la proposition principale de notre rapport, qui appelle à la fois détermination et raison.

On sait en effet que lorsqu'il s'agit de rémunérer des agents dont les tâches sont multiples, des indicateurs partiels conduisent à une focalisation excessive, l'agent délaissant celles qui ne sont pas comptabilisées. Il faut donc résister à la tentation de trop utiliser des indicateurs qui vont générer ce type de biais, en distinguant bien, par exemple, l'évaluation des performances des services et l'évaluation individuelle, et en s'attachant à prendre en compte dans ce cas ce qui contribue au potentiel de long terme. Par ailleurs, il faut construire l'acceptabilité d'un système d'indicateurs pertinents. Ceci nécessite de montrer aux agents qui seront jugés dessus qu'ils « ont leur chance », en clarifiant les objectifs fixés, en explicitant les solutions possibles permettant de les atteindre, et en fournissant l'accompagnement de formation correspondant. Ces précautions ne sauraient cependant justifier de différer l'élaboration de ces systèmes d'indicateurs, bien au contraire.

Recommandations

Des services et agences rendant compte des services rendus dans des conditions permettant la comparaison des performances...

1. Favoriser les bonnes pratiques suivantes (Fonds de réforme de l'État)

- Certification (type ISO 9000) des services ;
- Enquêtes auprès du public et des usagers ;
- Mesure de la qualité du service rendu ;
- Développement des approches en termes de *benchmarking* (méthodologies opérationnelles/applications) ;
- Certification de l'évaluation ;
- Soutien à la recherche en appui à la gestion publique.

2. Établir (circulaire du Premier ministre) le processus d' « *accountability* » au niveau des unités élémentaires de gestion (ou de fourniture du service)

- Poser que toute entité (école, hôpital, université, centre de recherche...) doit rendre compte à son autorité d'orientation (et en tout état de cause au public) du service rendu et de sa performance, sur des *bases permettant la comparaison* avec des services ou agences exerçant des activités similaires ;
- Que des rapports spécifiques doivent éventuellement compléter le rapport principal, en cas d'autorités multiples ;
- Que la mise en œuvre relève de la compétence de chaque ministre. À charge pour chacun d'eux de fournir des bases méthodologiques pour ce type de « *reporting* », utilisables dès l'exécution 2008.

3. Assurer la qualité des « agences d'évaluation », en développant des règles communes en matière de :

- Prévention des conflits d'intérêt,
- Certification externe des évaluations,
- Publicité et de transparence des évaluations,
- Renouvellement des comités, etc.

4. (Re)créer une capacité crédible à mener des études d'impact incluant de véritables comparaisons coûts-avantages des politiques publiques

- Préalable : établir la réalité actuelle de ces capacités (mission à confier à une personnalité faisant autorité en ce domaine) ;
- Stimuler la recherche appropriée pour fournir des références permettant d'évaluer tous les types de projets.

5. Systématiser les études d'impact des changements de formules de rémunération des agents ou des services, pour apprécier leur caractère incitatif, et identifier les éventuelles mesures d'accompagnement à mettre en place.

Références bibliographiques

- Arthuis J. (2005) : « Les objectifs et les indicateurs de performance de la LOLF », *Rapport d'Information du Sénat*, n° 220.
- Atkinson T. (2005) : *Atkinson Review: Final report, Measurement of Government Output and Productivity for the National Accounts*, Palgrave, McMillan, Londres.
- Beesley T. et M. Gathak (2003) : « Incentives, Choice and Accountability in the Provision of Public Services », *Oxford Review of Economic Policy*, n° 19(2), pp. 235-249.
- Bureau D. (2002) : « La santé à l'école de la régulation », *Societal*, n° 36.
- Burgess S. et M. Ratto (2003) : « The Role of Incentives in the Public Sector: Issues and Evidence », *Oxford Review of Economic Policy*, n° 19(2), pp. 285-300.
- Cahiers français (2005) : *La microéconomie en pratique*, n° 327, La Documentation française.
- Cohen E. (1997) : « Secteur public : portée et limites des pratiques de l'État actionnaire » in *Service public, secteur public*, Rapport du CAE, n° 3, La Documentation française.
- Cohen E. et C. Henry (1997) : « Sur les bases et l'évolution récente des services publics et industriels et commerciaux en France et dans l'Union européenne » in *Service public, secteur public*, Rapport du CAE, n° 3, La Documentation française.
- Bousquet F. et A. Coulomb (2007) : « Mesure de la performance et gestion hospitalière », Complément A in *Performance, incitations et gestion publique*, Rapport du CAE, n° 66, La Documentation française.
- Dewatripont M., I. Jewitt et J. Tirole (1999) : « The Economics of Career Concern, Part II: Application to Missions and Accountability of Government Agencies », *Review of Economic Studies*, n° 66, pp. 199-217.
- Dixit A. (1996) : *The Making of Economic Policy*, MIT Press, Cambridge, MA.
- Dixit A. (2002) : « Incentives and Organization in the Public Sector: An Interpretative Review », *Journal of Human Resources*, n° 37(4), pp. 696-727.
- Downs A. (1957) : *An Economic Theory of Democracy*, Harper and Row, New York.

- Gaffard J-L. (2007) : « L'évaluation de l'enseignement supérieur et de la recherche : le cas d'école de la réforme « LMD », Complément B in *Performance, incitations et gestion publique*, Rapport du CAE, n° 66, La Documentation française.
- Grout P.A. et M. Stevens (2003) : « The Assessment: Financing and Managing Public Services », *Oxford Review of Economic Policy*, n° 19(2), pp. 215-234.
- Guide méthodologique pour l'application de la loi organique relative aux lois de finances du 1^{er} août 2001 (2004) : *La démarche de performance : stratégie, objectifs, indicateurs*, La Documentation française.
- Hel-Thelier S. (2007) : « La mesure de la production et de la productivité des administrations publiques : la mission Atkinson au Royaume-Uni », Complément C in *Performance, incitations et gestion publique*, Rapport du CAE, n° 66, La Documentation française.
- Holmstrom B. (1982) : « Managerial Incentives Problems: A Dynamic Perspective », *Review of Economic Studies*, n° 66(1).
- Holmstrom B. et P. Milgrom (1986) : « Multi-Task Principal-Agent Analysis: Incentives Contracts, Asset Ownership and Job Design », *Journal of Law, Economics and Organization*, n° 7, 24-52.
- Institut de la gestion déléguée (2005) : *Quelle compétition pour l'amélioration du service public ?*, Rapport du groupe présidé par Ch. Babusiaux, La Documentation française.
- Kravchuk R. et R. Schack (1996) : « Designing Effective Performance-Measurement Systems Under the Government Performance and Results Act of 1993 », *Public Administration Review*, n° 56(4), pp. 348-358.
- Kruger F. (2002) : *La gestion des ressources humaines publiques à l'heure de la performance*, Institut de l'entreprise, Paris.
- Lacaze A. (2005) : « La LOLF : simple outil de management ou dogme écrasant ? Gérer et comprendre », *Annales des Mines*, n° 81, septembre.
- Laffont J-J. (1999) : « Étapes vers un État moderne : une analyse économique », in *État et gestion publique*, Rapport du CAE n° 24, La Documentation française, pp. 117-149.
- Laffont J-J. (2000) : *Incentives and Political Economy*, Oxford University Press.
- Laffont J-J. et J. Tirole (1986) : « Using Cost Observation to Regulate Firms », *Journal of Political Economy*, n° 94, pp. 614-641.
- Laffont J-J. et J. Tirole (1993) : *A Theory of Incentives in Procurement*, MIT Press.
- Lambert A. et D. Migaud (2005) : *Rapport au gouvernement sur la mise en place de la LOLF*.

- Manning W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz et M. Marquis (1987) : « Health Insurance and the Demand for Medical Care: Evidence From Randomised Experiment », *American Economic Review*, n° 773, juin.
- Milgrom P. et J. Roberts (1990) : « The Economics of Modern Manufacturing: Technology, Strategy and Organization », *American Economic Review*, n° 80 (3), pp. 511-528.
- Mougeot M. (1999) : *Régulation du système de santé*, Rapport du CAE, n° 13, La Documentation française.
- Mougeot M. et F. Naegelen (2005) : « La concurrence pour le marché », *Revue d'Économie Politique*, n° 115(6), pp. 739-778.
- OCDE (2000) : « Renforcer l'éthique dans le service public ; les mesures des pays de l'OCDE », *Note de Synthèse PUMA*, n° 7.
- OCDE (2005) : *Performance Related Pay Policies for Government Employees*.
- Perbet M. (2007) : « L'évaluation en France : retour d'expérience », Complément D in *Performance, incitations et gestion publique*, Rapport du CAE, n° 66, La Documentation française.
- Pochard M. (2003a) : *Rapport public du Conseil d'État*. La Documentation française.
- Pochard M. (2003b) : « Perspectives pour la fonction publique », *Futuribles*, n° 291, novembre.
- Propper C. et D. Wilson (2003) : The Use and Usefulness of Performance Measures in the Public Sector, *Oxford Review of Economic Policy*, n° 19(2), pp. 250-267.
- Sappington D.M. et J. Stiglitz (1987) : « Privatization, Information and Incentives », *Journal of Policy Analysis and Management*, n° 6, pp. 567-582.
- Tirole J. (1994) : « The Internal Organization of Government », *Oxford Economics Papers*, n° 46, pp. 1-29.
- Tirole J. (2005) : *The Theory of Corporate Finance*, Princeton University Press.
- Tobin J. (1970) : « On Limiting the Domain of Inequality », *Journal of Law and Economics*, n° 13, pp. 263-275.
- Wilson J.Q. (1989) : *Bureaucracy: What Government Agencies Do and Why They Do It*, Basic Books, New York.

Commentaire

Philippe Mongin

Directeur de recherche au CNRS et Professeur aux HEC

Le rapport se propose de faire avancer les deux questions, vivement débattues aujourd'hui, de *la performance de la gestion publique* et *des moyens incitatifs de l'améliorer*. À cette double fin, il convoque la théorie microéconomique des incitations, qui imprègne la plupart de ses observations positives et de son essai de recommandation. Plutôt que de suivre le texte dans son mouvement propre, on a choisi de contextualiser fortement le commentaire tout en le déportant quelque peu de l'objet initial. On rapprochera les questions traitées de la désormais célèbre loi organique de 2001, dite LOLF, qui a révolutionné la procédure et le contenu des lois budgétaires et qui, dans l'esprit de ses partisans, vise à soumettre le fonctionnement de l'État au concept nouveau de performance. Simultanément, on traitera de *politiques publiques*, suivant l'expression désormais reçue, qui est plus englobante que celle de *gestion publique*⁽¹⁾. Après avoir isolé ce qui concernait la performance dans les dispositions législatives, on ébauchera l'analyse du concept tout en veillant à préserver la terminologie concrète de la LOLF. On en viendra finalement à réapprécier l'apport des théories économiques ; en particulier, on insistera sur les limites de la théorie des incitations lorsque la fonction-objectif de la puissance publique est sous-définie.

(1) Les auteurs distinguent la gestion publique et privée des services publics et concentrent leurs réflexions sur la première forme. La notion de *politique publique* couvre tout ce qui tend à la réalisation des objectifs traditionnels de la puissance publique et englobe donc les deux formes de gestion. Elle provient des sciences politiques, mais elle a maintenant droit de cité dans la terminologie officielle de la République (la LOLF évoquée ci-dessous y fait référence).

1. La LOLF, occasion d'une remise en cause

La loi organique du 1^{er} août 2001 relative aux lois de finances, dite LOLF, modifie radicalement les principes et les procédures des finances publiques françaises, qui n'avaient pas sensiblement varié depuis les débuts de la V^e République. La nature de cette révolution mentale et technique a été souvent exposée par ses protagonistes, qui sont, d'une part, les parlementaires Alain Lambert et Didier Migaud, lesquels conçurent et défendirent la proposition de loi, et d'autre part, le ministre du Budget Jean-François Copé, auquel il revint d'en superviser la première application dans la loi de finances 2006. Un autre rapport du CAE, *Économie politique de la LOLF*, revient en détail sur les principes, qu'il n'y a donc pas lieu de récapituler ici⁽²⁾. On se contentera d'abstraire ceux qui régissent la définition des performances publiques, le choix de leurs mesures et les règles de leur constatation.

D'après les commentaires autorisés, la LOLF entend faire passer le budget de l'État d'une « logique de moyens » à une « logique de résultats », ou encore, suivant une autre expression révélatrice, d'instaurer pour les finances publiques un « pilotage d'après les performances ». L'article 51 dispose plus précisément que les actions de l'État doivent être mises au regard « des coûts associés, des objectifs poursuivis, des résultats obtenus et attendus pour les années à venir mesurés au moyen d'indicateurs précis dont le choix est justifié ». Cette logique comparative s'exprime dans le *Projet annuel de performance* (PAP) et le *Rapport annuel de performance* (RAP), qui articulent les objectifs et les indicateurs associés de deux manières complémentaires : dans le PAP, les indicateurs permettent d'appréhender les tendances et de poser les objectifs, qui prennent la forme de cibles différentielles par rapport à ces tendances ; dans le RAP, ils servent à rapprocher les cibles des résultats, donc à faire savoir dans quelle mesure les objectifs sont atteints. Comme le RAP est le double rétrospectif du PAP, les deux documents doivent s'appuyer sur les mêmes définitions des objectifs et des indicateurs. La LOLF incite à stabiliser ces définitions au-delà de l'année d'exercice, tant pour étendre les comparaisons que pour instaurer une planification pluriannuelle des dépenses.

Le législateur distingue hiérarchiquement *la mission*, qu'elle rattache à une « politique publique définie », *le programme*, unité intermédiaire que surplombe la mission et qui domine *l'action*, celle-ci représentant l'unité concrète de dépense publique. Le Parlement discute le budget mission par mission, sans pouvoir en créer de nouvelles, mais en modifiant s'il le souhaite les ressources allouées aux programmes au sein de l'une d'entre elles. Le programme est une unité d'exécution budgétaire confiée à un responsable, qui peut, le moment venu, modifier la répartition projetée des ressources entre les différentes actions. C'est au niveau intermédiaire des programmes que s'énoncent les objectifs et leurs indicateurs.

(2) Rapport de Edward Arkwright, Christian de Boissieu, Jean-Hervé Lorenzi et Julien Samson.

Dans l'esprit de la LOLF, les indicateurs revêtent forcément la forme numérique. Ils jouent un double rôle en liaison avec la dualité précédente du PAP et du RAP ; dans le détail, il apparaît que les comparaisons qui les impliquent sont loin d'être élémentaires. Les responsables s'engagent sur les cibles fixées dans le PAP et répondent de l'écart éventuel que constate le RAP. Ainsi, la LOLF a fixé des règles non seulement de l'évaluation, mais de la reddition de compte⁽³⁾, ce qui en fait le point d'ancrage éventuel d'un système incitatif de l'État et de ses agents. Le *Guide méthodologique pour l'application de la LOLF*, en date de juin 2004, dit aussi *Guide de la performance*, élabore ces considérations que l'on vient de résumer en les stylisant.

Un fragment de la loi de finances 2006 illustrera les concepts nouveaux. La mission « Recherche et enseignement supérieur » comporte, à côté d'autres, un programme « Formation supérieure et recherche universitaire », qui se répartit en différentes actions budgétaires. Le programme dans son ensemble se voit associer un objectif n° 1, « Répondre aux besoins de qualification supérieure », et un objectif n° 2, « Améliorer la réussite à tous les niveaux de formation ». L'objectif n° 1 donne lieu à deux indicateurs qui sont (i) le pourcentage d'une classe d'âge titulaire d'un diplôme de l'enseignement supérieur et (ii) l'insertion professionnelle des jeunes diplômés trois années après la formation initiale. C'est donc par rapport à ces deux données – la seconde étant moins immédiate à chiffrer que la première – que l'objectif s'exprimera par des cibles et que le législateur évaluera finalement le succès de sa réalisation. Un jugement sur l'objectif n° 1 ne mettrait pas un terme à la reddition de compte, puisque le responsable demande aussi à être jugé sur l'objectif n° 2. De même, l'indicateur (i) ne suffit pas en l'absence de l'indicateur (ii). À l'évidence, des difficultés de comparaison surgissent à ce point ; on les reprendra dans la section suivante.

Celle-ci commencera par signaler que la notion de performance est confuse dans le langage ordinaire ; ne serait-ce que pour cette raison, il était pertinent de rappeler comment les instigateurs de la LOLF s'y sont pris pour en déterminer mieux la signification. Le but du retour en arrière était aussi d'indiquer le contexte défini dans lequel les questions du rapport que l'on commente sont apparues. Les deux auteurs font état de la LOLF à la fois dans leurs motivations et dans certaines de leurs analyses ultérieures, mais – peut-être dans un souci louable de ne pas empiéter sur l'autre rapport du CAE – ils n'y touchent que d'assez loin, et c'est dommage pour l'utilisateur qui les attendait sur ce terrain d'application privilégié.

Si les auteurs jugeaient satisfaisante la notion concrète de performance qu'instaure la LOLF, ils pouvaient s'en servir comme du répondant concret

(3) Il est inutile d'emprunter le vocable anglo-saxon « *accountability* », alors que « reddition de compte » dit assez clairement ce qu'on veut dire. On veut en effet signifier à la fois l'existence d'une responsabilité personnelle de l'agent et le pouvoir qu'a l'autorité supérieure de mettre en cause cette responsabilité en réclamant des comptes.

de leur travail analytique. Si, au contraire, ils la jugeaient théoriquement inadaptée, ils pouvaient le faire savoir en développant leurs raisons. À les lire entre les lignes, on conclura qu'ils jugent sévèrement les indicateurs de la LOLF, d'abord au motif qu'ils ne seraient pas tous bien choisis (alors que la loi stipule qu'ils doivent être « pertinents », « utiles », « solides », « vérifiables »), ensuite pour la raison qu'ils seraient exagérément « globaux » et, par là, donneraient une image douteuse de la performance. La première critique est évidente à faire, mais elle peut très bien ne refléter qu'un état de choses temporaire. La lecture de la loi de finances 2006 révèle des faiblesses dans le choix des indicateurs qui seraient faciles à rectifier. Plus substantielle bien qu'obscur, la seconde critique se rapproche d'une objection véritable contre le dispositif nouveau. Il est en effet concevable que celui-ci échoue non seulement parce qu'il devrait conduire à des sanctions et des incitations qui ne sont pas encore définies, mais aussi et plus gravement, parce qu'il ne réussirait pas non plus sur le terrain limité de l'évaluation rétrospective.

La loi de 2001 a la particularité d'instaurer l'évaluation des performances tout en liant son déroulement à l'exercice contraint de l'activité budgétaire. Cette liaison est à la fois une chance et un danger pour la réforme de l'État : une chance, parce qu'on peut ainsi espérer que le travail sur la performance étatique ne tombera pas dans l'insignifiance des essais précédents, comme la rationalisation des choix budgétaires (RCB), dont l'une des annexes retrace utilement l'échec ; mais un danger aussi, parce que les concepts et les mesures se ressentent, dans leur usage et même dans leur définition, de la structure bureaucratique à laquelle ils sont finalement destinés. C'est donc une thèse à première vue défendable que la LOLF serait mal adaptée à la réforme de l'État qu'elle cherche à préparer ; encore faudrait-il que cette thèse soit effectivement défendue, et non pas distillée sous forme d'allusions.

2. Qu'est-ce que la performance étatique ?

Le langage ordinaire traite les mots « performances » et « résultats » comme plus ou moins interchangeables. Il en va d'ailleurs ainsi dans certains textes à caractère économique : les notes de conjoncture parlent des performances de l'Allemagne à l'exportation presque dans le même sens que *L'Équipe* des performances de Lance Armstrong au Tour de France. On se fixe sur les réalisations apparentes du cycliste ou du pays en sous-entendant qu'elles lui seraient imputables, et tout naturellement, on passe aux jugements qui approuvent ou désapprouvent. L'Allemagne « fait bien » à l'exportation même si les bons chiffres s'expliquent aussi par la chute momentanée de la demande intérieure ou par un effet-prix anormal dû à la montée de l'euro. Lance Armstrong a « fait un sans-faute » au dernier Tour même si, à ce qu'il paraît, des adjuvants ne seraient pas sans rapport avec le triomphe.

Il va de soi que la performance, en matière de politiques publiques, doit se définir plus abstraitement. On veut identifier ce qui, dans le résultat, dépend exclusivement de la politique envisagée ; il convient d'écarter par la pensée – et, si possible, par la mesure – tout ce qui s'impute à d'autres facteurs causaux. Ainsi, l'objectif « Répondre aux besoins de qualification supérieure » peut s'atteindre différemment selon que le marché de l'emploi est florissant ou déprimé. Le placement des tourneurs-fraiseurs à la sortie des BTS peut être de 95 % non parce qu'ils seraient bien formés pour leur tâche, mais simplement parce que l'industrie mécanique se trouve en phase haute et qu'elle se rabat sur le tout-venant. De manière générale, il n'y a pas lieu d'invoquer la sanction du marché si l'on ne sait pas distinguer les contributions respectives de l'offre et de la demande à la réalisation de l'objectif.

Le problème d'identification n'est pas seulement lié à la présence de marchés, comme le montrent les exemples médicaux que cite le rapport et que développe l'un de ses compléments. Les programmes visant à prévenir le cancer du sein ou à diminuer les maladies nosocomiales ont des résultats qui dépendent bien sûr d'autres facteurs causaux que leur simple mise en œuvre ; il s'agit par exemple, dans le premier cas, de la bonne information du public féminin, et dans le second cas, de l'intériorisation des normes d'hygiène par les personnels hospitaliers. Certes, les facteurs supplémentaires ne sont pas toujours exogènes aux politiques, loin s'en faut, mais s'ils font l'objet d'une intervention, celle-ci ne relève pas nécessairement du programme dont on cherche à identifier la performance. Sans doute la lutte contre les maladies nosocomiales peut-elle s'adjoindre les campagnes de persuasion destinées aux personnels, de sorte que, dans ce cas, le même programme prendra en charge à la fois l'action principale et le facteur supplémentaire. Mais la prévention du cancer du sein justifie probablement de séparer le programme qui rend les tests disponibles et ceux qui visent à informer les utilisatrices – on s'en remettra au planning familial et à la médecine du travail, on lancera des campagnes d'information préventive qui couvrent aussi d'autres cancers, et ainsi de suite. Il arrivera donc que plusieurs programmes – voire plusieurs missions ou politiques publiques – collaborent pour atteindre un même résultat, ce qui complique singulièrement l'imputation des performances.

De telles difficultés appellent des solutions pragmatiques dont les auteurs pouvaient faire le bilan. Ils identifient une méthode possible lorsqu'ils écrivent que la performance doit s'estimer « en moyenne ». Mais sauf à préciser les moyens économétriques de la raffiner, la réponse ne convient que si les facteurs qui influencent le résultat en sus du programme qu'on évalue satisfont aux propriétés statistiques ordinaires, et tout particulièrement, sont indépendants en probabilité des actions couvertes par ce programme. En toute rigueur, la règle demande plus encore : il faudrait que ces autres facteurs soient *exogènes à la totalité de l'action de l'État*. Que dirait-on d'une appréciation « en moyenne » de la mise à disposition de tests médi-

caux qui ne séparerait pas les périodes où le public était informé de celles où il ne l'était pas⁽⁴⁾ ?

Pour limiter le travail statistique considérable qui se profile, il serait concevable de *redéfinir* les programmes à évaluer, soit en regroupant ceux qui collaborent à un même résultat, soit en subdivisant à la fois les programmes et les résultats de manière à mettre les uns et les autres en relation biunivoque. Tous les regroupements sont réalisables, mais toutes les subdivisions ne le sont pas. Cette dissymétrie tient au fait qu'un programme peut toujours s'élargir à d'autres actions, alors que les actions qui lui servent de constituants s'avèrent souvent indivisibles et multifonctionnelles. Ainsi, la mise à quatre voies d'une route nationale obéit aux objectifs simultanés de faire baisser le nombre d'accidents qui s'y produisent, d'améliorer la desserte des villes qu'elle traverse et de faire gagner du temps et du carburant aux voyageurs qui l'empruntent. Il est matériellement irréalisable de séparer trois sous-actions qui refléteraient chacune un des trois objectifs. En revanche, rien n'empêche de définir un programme d'amélioration routière tellement vaste qu'il couvrirait toutes les actions destinées à faire baisser le nombre d'accidents, toutes les actions qui améliorent les dessertes locales, toutes les actions qui font économiser du temps et du carburant sur les trajets longs.

Mais si l'on prend cette direction généralisante, on butte sur des objections immédiates, les unes normatives et les autres simplement factuelles. En premier lieu, un programme ne doit pas être si largement conçu que ses objectifs deviennent trop vagues ou trop éclatés, et il ne faut pas non plus que son responsable dispose de tellement d'actions possibles qu'il échappe au contrôle du supérieur ou de l'autorité politique ; et l'on ne parlera même pas des pertes d'efficacité qui accompagnent banalement le gigantisme. En second lieu, c'est un fait que l'État et les autres institutions publiques restent structurés en unités de moyens plutôt que de fins. La multiplication récente des ministères qu'on peut dire programmatiques – environnement, ville, condition féminine, égalité des chances – ne compensera pas de sitôt la prédominance traditionnelle des administrations qui se caractérisent par les moyens dont elles disposent. Plus l'objectif d'un programme s'élargit, plus nombreux seront les bureaux dont sa réalisation dépend, et l'on peut craindre qu'à force de redéfinitions, il ne requière l'appareil entier de l'État. En bref, il apparaît que, si l'on facilite d'un côté l'appréciation des performances en fusionnant les programmes, on la complique d'un autre côté en rompant le lien que celles-ci entretiennent avec des services administratifs désignés.

(4) Ces remarques triviales ne sont pas à la mesure des vastes problèmes statistiques et économétriques soulevés par l'évaluation des politiques publiques. Le sujet a fait l'objet de synthèses récentes, notamment par Thierry Magnac (2000) : « L'apport de la microéconométrie à l'évaluation des politiques publiques », *Cahiers d'économie et de sociologie rurale*, et par Thomas Brodaty, Bruno Crépon et Denis Fougère (2007) : « Les méthodes micro-économétriques d'évaluation : développements récents et applications aux politiques actives de l'emploi », *Économie et prévision*, à paraître.

La définition des indicateurs apporte son lot de difficultés supplémentaires. Le rapport souligne à juste titre qu'elle peut induire, chez les responsables de programme, certains comportements décalés ou même dissociés de la réalisation des objectifs que ces indicateurs ont pour but d'exprimer. Les effets pervers de ce genre tombent sous deux catégories. Il se peut tout d'abord que l'objectif ne se mesure qu'en partie et que les responsables se fixent sur les actions qui ont des *conséquences mesurables* aux dépens de celles qui n'en ont pas, mais qui feraient également progresser l'objectif. Mais il se peut aussi que les valeurs prises par l'indicateur fassent l'objet d'une *manipulation directe ou indirecte*, de sorte qu'elles n'indiquent plus grand-chose de l'objectif recherché. En prenant la métaphore visuelle, on dira que, dans le premier cas, l'image de la performance est nette, mais incorrecte, alors que, dans le second cas, les contours de cette image se brouillent au point que l'objet n'est plus reconnaissable. Le rapport cite les exemples très pertinents de la tarification à l'acte, qui est censée améliorer la gestion des hôpitaux, et des statistiques d'activités délictueuses, qui, avec la LOLF, servent de mesures de performance à la police et à la gendarmerie.

Voici un autre exemple, qui a l'avantage d'illustrer le glissement toujours possible du premier genre de distorsion vers le second. Par sa nature même, l'objectif qui consiste à « Répondre aux besoins de qualification supérieure » ne peut donner lieu qu'à des indicateurs chiffrés partiels et discutables. Encore y a-t-il lieu de ne pas mettre sur le même plan des tests de qualification qui seraient confiés à des jurys extérieurs et l'indicateur (i) que la loi de finances a malheureusement retenu : « le pourcentage d'une classe d'âge titulaire d'un diplôme de l'enseignement supérieur ». Il ne faut pas connaître beaucoup les universités pour conclure que cet indicateur est manipulable. Le ministère de l'Éducation nationale ne peut pas ignorer qu'en liant la répartition des ressources universitaires à la réussite aux examens, il encourage la bienveillance intéressée des enseignants qui dirigent les filières. Les Présidents d'université retransmettront s'il le faut la consigne discrète ; ainsi, la chaîne de dépendance fait que l'indicateur devient manipulable par le responsable de programme lui-même. La critique se transpose à d'autres valeurs-cibles, comme les trop célèbres « 80 % d'une classe d'âge au baccalauréat » : cet indicateur ne peut pas refléter précisément l'objectif d'élévation du niveau scolaire puisque les enseignants du secondaire, intéressés à la réussite, se voient confier la correction de l'examen. Dans le domaine de l'enseignement plus qu'ailleurs, la logique de la performance risque d'être compromise par des distorsions des deux types. Le complément consacré à la réforme « LMD » confirme par d'autres moyens ce diagnostic sévère. L'Éducation nationale pourrait s'imposer la contrainte modeste d'évacuer les indicateurs qui s'offrent aux distorsions du second type, sachant que celles du premier type sont beaucoup plus complexes à maîtriser et qu'il en subsistera inéluctablement.

Les distorsions traduisent un problème incitatif fondamental que le rapport fait bien de souligner. Mais il faut immédiatement ajouter que ce pro-

blème n'est ni le seul que soulèvent les indicateurs, ni même celui qu'ils posent en premier lieu à l'analyse normative. Car celle-ci doit commencer par réduire la multiplicité des indicateurs qui représentent un même objectif et la multiplicité concomitante des objectifs qui dirigent une même politique publique. La double prolifération risque d'interdire le jugement attendu sur les performances, *et cela indépendamment de l'existence ou non de perturbations incitatives*. Supposons, pour les besoins de la démonstration, que les indicateurs couvrent les principaux aspects de l'objectif et qu'ils ne se laissent pas manipuler, même indirectement ; il resterait encore à les synthétiser et, idéalement, à les ramener à une dimension scalaire unique. Comment, par exemple rapprocher (i) de (ii), « l'insertion professionnelle des jeunes diplômés trois années après la formation initiale » ? Si l'on ne rend pas commensurables ces deux indicateurs, on ne pourra pas évaluer la réalisation de l'objectif qu'ils expriment censément, « Répondre aux besoins de qualification supérieure ». De même, il convient de synthétiser les objectifs multiples qui correspondent à une politique publique donnée. La méthode à suivre n'est pas exactement la même que pour les indicateurs, parce que ceux-ci ont une forme quantitative dont les objectifs sont en général dépourvus, et parce qu'il faut, cette fois, opérer des arbitrages substantiels, et non pas seulement condenser des informations disparates.

Si le rapport ne traite pas ces problèmes, ce n'est sans doute pas qu'il les juge négligeables, mais c'est plutôt parce qu'il se place d'emblée au point de vue théorique des incitations. Il n'est pas incorrect de procéder ainsi tant qu'on se propose uniquement de faire l'analyse positive du fonctionnement bureaucratique. De fait, il n'y a pas lieu de synthétiser les indicateurs multiples si, justement, on se propose de décrire les effets incitatifs pervers de cette multiplicité. À ce point de vue, Dominique Bureau et Michel Mougeot font un travail décapant qui rendra service aux fonctionnaires et aux parlementaires qui, dans les années à venir, prépareront les lois de finances en conformité avec les exigences de la LOLF. Mais si le but final du rapport était de *recommander* un système incitatif, il lui manque quelque chose, puisqu'il faut commencer par aider la puissance publique à clarifier ce qu'elle veut inciter ses agents à réaliser. Les considérations précédentes deviennent alors inéluctables.

3. La pertinence des théories économiques

La critique faite en dernier lieu donne à penser qu'à côté de la théorie des incitations, il y avait place, dans ce rapport, pour d'autres théories économiques : *la théorie de la décision*, qui formalise l'agrégation des données numériques multiples, et *l'économie du bien-être* (ou, sous sa forme postérieure, *la théorie du choix social*), qui met en place les « fonctions de bien-être social » dont la définition précède logiquement le problème normatif de l'incitation étatique. Le point de vue du bien-être social aurait eu l'avan-

tage supplémentaire de faire mieux voir que les politiques publiques ne se jugent pas seulement à l'aune de leurs résultats désirables, mais encore à celle de leurs conséquences nuisibles. La mise à quatre voies d'une route nationale fait une saignée dans le paysage, crée des nuisances sonores, dévalorise les maisons mitoyennes. La création de diplômes qui répondent à des besoins socio-économiques déterminés peut détourner les étudiants des matières fondamentales et nuire ainsi à la recherche scientifique. En bref, il n'est guère de dépenses publiques dont l'engagement ne comporte aussi une *externalité négative*, de sorte que l'évaluation se présente d'emblée comme un problème multidimensionnel : ce n'est pas seulement qu'on doive arbitrer entre plusieurs objectifs, c'est qu'il faut en outre mettre ces objectifs en balance avec les coûts dont s'accompagne leur réalisation. Le rapport n'entre pas dans ces considérations qui, on doit l'accorder, cheminent lentement chez les économistes en général ; il est aussi regrettable qu'elles soient restées étrangères à l'inspiration de la LOLF.

Quant à la théorie des incitations, elle couvre aujourd'hui des développements nombreux, mais peu coordonnés, de sorte que certains de ses praticiens ne diraient pas qu'il s'agisse une théorie en bonne et due forme. En tout cas, ce n'est pas une théorie *intégrée* comme le sont – avec tous les risques d'étroitesse et de répétition que ce caractère implique – les deux théories qui viennent d'être appelées maintenant à la rescousse. Le rapport adopte un ton de généralité que ne justifie pas encore un courant qui excelle surtout dans la résolution habile de cas particuliers successifs. Au-delà de l'apparence, l'éclatement des modèles est perceptible dès qu'on prend la peine de consulter directement les travaux – toujours bien choisis – que citent Dominique Bureau et Michel Mougeot ; et s'ils peinent finalement à livrer des recommandations précises, c'est aussi bien que la réalité de ces travaux reprend sévèrement ses droits.

Les auteurs s'attachent à la variante du problème incitatif que formalise la notion microéconomique de *contrat* : un principal confie une tâche à un agent qu'il rémunère en fonction de ce qu'il observe de l'accomplissement de cette tâche. Une structure aussi abstraite peut se transposer de multiples manières aux administrations publiques. L'application la plus brutale concernerait ces administrations prises toutes ensemble quand elles font face aux autorités politiques : celles-ci représentent le principal, et celles-là, les agents. Avec plus de subtilité, on identifie des relations principal-agent qui lient entre eux les niveaux hiérarchiques des administrations centrales, les services centraux et leurs services déconcentrés, les autorités de tutelle et leurs institutions subordonnées, les directeurs d'établissement et leurs employés, l'administration et ses fournisseurs. Les applications que le rapport envisage se situent à ces niveaux intermédiaires, et comme telles, restent parlantes et défendables ; l'exemple déjà cité des normes hospitalières est même spécialement pertinent. Les auteurs font grand cas de la distinction classique entre deux types d'information incomplète, celle du type « antisélection » (l'agent connaît mieux les états du monde que le principal)

et celle du type « aléa moral » (l'agent peut accomplir certaines actions à l'insu du principal). Déjà inscrite dans le sens commun, cette distinction est très utile ; on l'a d'ailleurs pratiquée plus haut lorsqu'on traitait du problème incitatif des indicateurs. Systématisée par la littérature des contrats, elle fait comprendre que la performance est affaire, non pas seulement de *culture*, mais encore de *structure d'information*, ce qui doit tempérer quelque peu l'enthousiasme soulevé par la LOLF.

Mais cette application décapante ne va guère plus loin que l'emploi des définitions premières. La microéconomie des contrats manifeste là clairement ses limites d'emploi. On a dit ce qu'il en était de ses résultats généraux. Quant à ses modèles particuliers, il apparaît qu'aucun d'entre eux ne réunit encore les hypothèses convenables pour traiter de performances publiques. Le modèle de base comporte un principal unique et des agents qui accomplissent une tâche exclusive en échange d'une rémunération qui dépend ponctuellement du travail qu'ils y consacrent. Ce modèle n'approfondit pas les objectifs du principal, soit qu'il les ramène au profit monétaire, soit qu'il les prenne plus abstraitement comme des données. Les hypothèses qu'on attendrait sont tout autres :

- des principaux multiples et liés entre eux par une relation d'autorité ;
- des agents qui accomplissent des tâches non exclusives et dont la rémunération peut revêtir des formes non monétaires ou décalées dans le temps ;
- des objectifs soigneusement identifiés qu'une fonction de bien-être social devra synthétiser quand l'analyse prendra un tour normatif.

Certains modèles avancés tirent dans le sens de (i) et de (ii), mais pour traiter ces complications une à une, sans les rendre pour l'instant concilia- bles, et ils négligent de toute façon la complication (iii), sur laquelle ce commentaire s'est plus particulièrement étendu.

Complément A

Penser une nouvelle place pour la qualité dans la régulation du système de santé français

Alain Coulomb et Frédéric Bousquet

Haute Autorité de santé

Ce complément fait suite à l'audition du directeur de la Haute Autorité de santé par le Conseil d'analyse économique qui souhaitait disposer de son expertise sur l'analyse comparée des performances des secteurs publics et privés dans le domaine hospitalier. Mettant en exergue les limites d'une approche exclusivement centrée sur la performance économique, il est apparu intéressant d'ouvrir le champ de la réflexion en soulignant l'importance, particulièrement dans le domaine de la santé, d'intégrer la qualité comme une dimension essentielle de la mesure de la performance, et dès lors, de faire de la qualité un axe fort des politiques de régulation. Ce faisant, on verra comment la qualité vient enrichir, de manière cohérente, les deux grands axes traditionnels de l'évaluation des politiques publiques en santé que sont l'efficacité, aussi bien médicale qu'économique, et l'équité.

La question de la mesure de la performance est entrée aujourd'hui dans les esprits de tous les dirigeants d'organisations publiques ou privées. Il serait toutefois prématuré de dire que la mesure elle-même est entrée dans les mœurs ou dans les cultures, et bien sûr dans les faits, qu'il s'agisse de l'entreprise ou du service public. Pourtant, l'idée fait son chemin et les pratiques évoluent, comme en témoigne la mise en œuvre de la loi organique relative aux lois de finances (LOLF).

Le concept même de performance, et la question de sa mesure, semble encore trop souvent rester enclavés dans un débat idéologique, parfois stérile. Le thème de la comparaison de la performance prête en effet à un

triple malentendu qui tient notamment au fait que ce concept est trop souvent limité à la notion de productivité.

Le premier est lié à l'idée trop souvent admise que la performance est une conception issue et propre au secteur de l'économie marchande qui montrerait ainsi la nette supériorité de son efficacité face à un secteur public qui lui préférerait une opacité plus tranquillisante. Or, il est intéressant de noter qu'outre-Manche et outre-Atlantique, la mesure de la performance, dans sa conception moderne, est apparue à peu près en même temps, dans le courant des années quatre-vingt, au sein des deux secteurs, constituant en cela un fait assez peu courant en matière d'évolution du management

Deuxième malentendu et corollaire direct du précédent, la mesure de la performance traduirait une forme de prosélytisme « néolibéral » du secteur marchand, et toute introduction de la notion de performance, en particulier économique, constituerait une vision à courte vue des services publics et une forme de privatisation larvée de ceux-ci. Certes, les missions de service public appellent sans doute à introduire dans les dimensions de la performance des concepts tels que les services rendus aux usagers et aux populations, mais les notions d'efficience ou de responsabilisation des producteurs publics envers la collectivité ne peuvent être occultées du débat démocratique.

Le troisième et dernier malentendu constitue souvent une forme de réaction de la part du secteur public à une vision parfois trop caricaturale de la performance économique. Il consiste en un rejet systématique de la validité et de l'application des concepts et des outils de mesures et conduit au déni des résultats fournis par les indicateurs et les comparaisons que l'on en tire. Ceci souligne sans doute l'impérative nécessité de mettre en place des outils de mesure d'une très grande rigueur méthodologique. À cet égard, les expériences montrent un très net progrès sur ce plan dans le secteur de la santé et il est aujourd'hui reconnu possible de fournir des mesures de la performance des établissements de santé.

Ce triple constat conduit aujourd'hui à repenser la place de la qualité qui représente une des dimensions essentielles de la performance dans la santé. Les progrès récents sur la mesure de la qualité ouvrent un nouveau paradigme en matière d'évaluation de la performance du système, qui se limitait traditionnellement aux axes de l'efficacité et l'équité. L'introduction de la qualité comme une variable endogène de l'action publique permet en effet de renforcer à la fois l'efficacité et l'équité du système. Toutefois, il reste encore à inventer de nouvelles modalités d'action des pouvoirs publics. À l'instar des exemples étrangers, il apparaît aujourd'hui souhaitable de développer des mécanismes incitatifs centrés sur la qualité.

1. Renforcer la culture de l'évaluation et de la mesure chez les professionnels

Le fait de mesurer la qualité a un impact direct sur le niveau de la qualité des soins. Son intérêt est double : il permet non seulement de faire émerger les défauts de qualité mais constitue également, dans bien des cas, un premier pas « spontané » vers la mise en œuvre d'actions correctrices ou d'amélioration.

L'existence d'un déficit de qualité dans nos systèmes de santé modernes n'a été constatée que récemment, essentiellement sur la base de données anglo-saxonnes. L'équipe de McGlynn a procédé à l'analyse de près de 7 000 dossiers concernant trente pathologies aiguës et chroniques ; plus de 400 indicateurs ont été mobilisés. Le résultat est édifiant : seulement 55 % des patients ont reçu des soins conformes à l'état de la science. En France, l'étude nationale sur les événements indésirables graves liés aux soins (ENEIS)⁽¹⁾, réalisée par le CCECQA en collaboration avec la DREES, a permis de montrer que 3 à 5 % des hospitalisations sont causées par un événement indésirable grave (EIG), soit 175 000 à 200 000 séjours, dont 70 000 à 110 000 pourraient être évités ; près de la moitié des EIG à l'origine d'une hospitalisation sont associés à des produits de santé.

Diverses expériences pilotes d'évaluation interne des pratiques professionnelles menées par des hôpitaux ou des cliniques ont montré que le fait même de mesurer est très souvent un premier pas vers une amélioration de la qualité : la phase de mise en œuvre de l'évaluation est toujours l'occasion de mieux formaliser les processus de soins, elle constitue souvent une prise de conscience des problèmes individuels ou organisationnels dans le fonctionnement d'une équipe, etc. Cependant, pour certains observateurs, ces expériences ne constituent pas une preuve de l'amélioration de la qualité du système de santé dans sa globalité : seules les équipes les plus motivées mettent en place des démarches d'amélioration, ces expériences reposent souvent sur la bonne volonté d'un petit nombre de professionnels dont la motivation peut s'épuiser au fil du temps. Aussi, l'enjeu n'est-il pas seulement de convaincre de l'efficacité ponctuelle des démarches qualité, mais surtout de les systématiser et les pérenniser.

(1) DREES (2005) : « Les événements indésirables graves liés aux soins », *Études et Résultats*, n° 398, mai.

2. Renforcer l'efficacité économique

Traditionnellement, l'économie de la santé appréhende la performance des systèmes de santé selon deux dimensions : leur efficacité et leur équité. La notion d'efficacité appliquée au secteur de la santé distingue en général l'efficacité médicale de l'efficacité économique. Cette dernière peut à son tour être déclinée d'une part en efficacité allocative (qui met en relation les coûts et la production et qui pose la question : les ressources allouées n'auraient-elles pas permis d'atteindre un niveau de satisfaction plus élevé si elles avaient été employées à un autre usage ?) et en efficacité productive (qui met en relation l'allocation des ressources et la production et qui pose la question : compte tenu des ressources allouées, obtient-on la production maximale ?).

2.1. La qualité peut souvent constituer un facteur de diminution des coûts de production

Les travaux méthodologiques sur les coûts de la qualité et leur application opérationnelle dans le secteur industriel ont montré que les démarches qualité permettaient d'abaisser les coûts de production. Afin de comprendre et de cerner les liens qui pourraient exister entre coûts et qualité dans le domaine de la santé, l'Agence nationale d'accréditation et d'évaluation en santé (ANAES) a mené une étude orientée vers le développement d'outils pragmatiques⁽²⁾ dont voici les principaux résultats.

Dans une approche conceptuelle, le « coût de la qualité » peut se décomposer en coûts d'investissement dans la qualité d'une part (i.e. les coûts de prévention des défauts de qualité correspondant à la formation du personnel, la mise en place d'équipes qualité, etc.) et coûts de mesure de la qualité d'autre part (études, analyses et matériels mis en œuvre). Ces coûts d'investissements doivent être mis en balance avec les coûts potentiels des défaillances ou des défauts de qualité⁽³⁾, ainsi que les coûts de leurs corrections par rapport au « surcoût » qu'elles ont permis d'éviter.

Ces concepts sont malheureusement rarement mobilisés dans le cadre d'évaluations. L'étude menée par l'ANAES, faute d'aboutir à un chiffrage global, a cependant permis de mieux cerner les défauts de qualité potentiellement très coûteux. Quatre points saillants ont ainsi pu être mis en évidence :

- une typologie des défauts de qualité les plus fréquents : événements indésirables, prescriptions inappropriées, interventions et hospitalisations non

(2) ANAES (2004) : *Les coûts de la qualité et de la non-qualité des soins dans les établissements de santé : état des lieux et propositions*, La Documentation française.

(3) Les défauts de qualité des soins sont habituellement classés en trois catégories : la sous-utilisation des ressources que l'on peut traduire par un déficit de soins, la surutilisation des ressources (hospitalisations inappropriées, prescriptions d'antibiotiques dans le cadre d'affections virales, etc.) et les défauts de réalisation d'un soin adéquat entraînant un événement indésirable évitable.

justifiées, défauts de circuit du médicament, non-adéquation des patients aux lits d'hospitalisation constituent des défauts de qualité qui peuvent être potentiellement très coûteux du fait de leur fréquence ;

- l'identification des niches coûteuses de non-qualité : les hospitalisations et interventions non justifiées, ainsi que les événements indésirables ressortent plus particulièrement des données françaises ;

- un ciblage plus précis des événements indésirables coûteux à partir des données étrangères : infections et complications post-opératoires infectieuses, bactériémies et pneumopathies, certains événements indésirables médicamenteux ; les hospitalisations et prescriptions inappropriées apparaissent aussi comme des défauts très coûteux à l'étranger ;

- un chiffrage non représentatif sur quelques exemples (en général des études monocentriques). En France, les rares études mobilisant des données de coûts abordent le sujet des infections nosocomiales. Un exemple : on estime à près de 400 euros en moyenne le coût des infections urinaires, celui des septicémies survenues en service de réanimation se situerait entre 26 000 et 35 000 euros.

2.2. La qualité peut-elle participer à une amélioration de l'allocation de ressources aux producteurs ?

Cette question recouvre en fait deux problématiques. La première est de savoir s'il existe un lien clair et simple entre niveau de qualité et niveau de ressources. La seconde pose plutôt la question de savoir si la création d'un lien entre niveau de rémunération et niveau de qualité conduit ou non au renforcement de l'efficacité du système de santé.

Sur le premier point, le simple bon sens plaide en faveur d'un lien qui, selon certains, serait quasiment proportionnel. Pourtant, il ne semble pas exister de corrélation entre les disparités des ressources des établissements de santé et les variations du niveau de qualité. À titre d'exemple, la Haute Autorité de santé a analysé les résultats de l'accréditation sur un échantillon d'établissements et tenté d'identifier les facteurs explicatifs des différences de qualité constatées, mesurées par la formulation ou non de recommandations ou de réserves à l'issue de la procédure. Ce critère d'évaluation de la qualité a ensuite été croisé avec la valeur du point ISA de chacun des établissements. Aucune différence de valeur du point n'a pu être mise en évidence entre les établissements accrédités sans réserve ni recommandation et ceux en ayant au moins une. Cet exemple ne constitue en rien une preuve scientifique, mais fournit toutefois de sérieux indices.

S'il est clairement démontré que les modes de financement affectent les comportements des prestataires de soins, rien ne permet de conclure aujourd'hui que la qualité représente pour ceux-ci une variable d'ajustement essentielle aux fluctuations du niveau de leurs ressources, dans un sens ou dans l'autre. Plusieurs travaux récents montrent en effet que des facteurs

de nature professionnelle (éthique, déontologie), sociale, psychologique ou morale peuvent très largement contrebalancer les effets d'une rationalité économique pure⁽⁴⁾.

Si l'on dispose de peu de données permettant de fixer un cadre conceptuel précis entre qualité et ressources, il existe cependant des indices forts suggérant qu'un accroissement de ressources ne se traduit pas nécessairement par un gain de qualité. Ces constats ne permettent toutefois pas d'épuiser la réflexion sur la deuxième question relative à la rémunération de la qualité.

Les pouvoirs publics seraient-ils plus « justes » en rémunérant mieux les professionnels reconnus comme fournissant une meilleure qualité de prestation ? Cela permettrait-il de renforcer la qualité globale du système de santé ? Ce faisant, ne risque-t-on pas de créer des effets pervers en incitant par exemple les professionnels à « tricher » sur leur niveau de qualité ? Il n'existe pas à ce jour de réponse définitive et « scientifique » à ces questions qui font seulement l'objet d'une exploration récente outre-Atlantique. Néanmoins, les principes guidant la réflexion dans ce domaine doivent conduire à éviter les mécanismes qui gèlent les situations par la simple reconnaissance, à travers une différence de rémunération, des « bons et des mauvais élèves », et à mettre en place au contraire les éléments qui permettront aux moins bons de rattraper les meilleurs. Il s'agit donc moins de récompenser ou de sanctionner que d'inciter. On reviendra sur ce point dans la troisième partie.

3. Qualité et équité : un choix de société

Le rapprochement des notions de qualité et d'équité renvoie bien évidemment à la problématique classique de l'accessibilité aux soins. La question qui se pose aux pouvoirs publics est donc de savoir comment garantir un égal accès à des soins de qualité identique en tout point du territoire. Elle ouvre dès lors la perspective de la mise en œuvre d'une politique de régulation de la qualité.

Les relations entre qualité et équité peuvent être appréhendées de deux manières. La première consiste à rendre les professionnels et les organisations plus « également efficaces » afin d'assurer à tous les citoyens un accès mieux partagé à des soins de meilleure qualité. À cette fin, différents outils ont été introduits : normes, références médicales opposables (RMO), recommandations, FMC, accords de bon usage des soins (AcBUS), accréditation des établissements, puis évaluation des pratiques professionnelles (EPP) et accréditation des médecins dans le cadre des activités à risque.

La deuxième manière d'aborder le sujet consiste à reconnaître l'existence d'un « droit de regard » des citoyens sur la qualité des prestations qui

(4) On pourra voir par exemple : Town et al. (2004) : « Assessing the Influence of Incentives on Physicians and Medical Groups », *Med. Care Res. Rev.*, septembre.

leur sont fournies. Cette deuxième conception implique ainsi les professionnels de santé dans une nouvelle exigence collective que les pays anglosaxons désignent sous le nom de « *accountability* » et que l'on pourrait traduire par la notion de responsabilité⁽⁵⁾. L'activité soignante relevant quasi exclusivement de la sphère collective, il s'agit pour ses acteurs de rendre compte de ce qu'ils font. La relation entre qualité et équité passe ici par l'obligation d'information des citoyens dans une logique d'exigence de démocratie sanitaire.

En France, les pouvoirs publics ont plutôt choisi la voie de l'auto-évaluation et de l'autorégulation des professionnels. Ce choix évolue toutefois rapidement en ce qui concerne les établissements de santé, comme en témoignent les demandes des ministres de la santé successifs de mettre en place des « palmarès » d'établissements. Il serait toutefois trop rapide de conclure que ces derniers fournissent la solution « clefs en main » aux attentes des usagers-citoyens. D'une part, les palmarès sont des outils méthodologiquement peu fiables apportant finalement une information peu utilisable. Par ailleurs, en stigmatisant, parfois injustement, certains établissements, ils concourent à opposer professionnels et usagers alors même que l'on cherche à construire un dialogue.

4. Les incitations, une nouvelle forme d'action des pouvoirs publics

4.1. Une base commune : un système de mesure de la performance

Aujourd'hui, les principaux systèmes de mesure de la performance hospitalière ont tous développé une structure qui s'appuie principalement sur le développement d'indicateurs. On distingue quatre phases :

- définir les axes prioritaires sur lesquels on souhaite développer des mesures et faire porter les efforts d'amélioration ;
- identifier et sélectionner les indicateurs dont l'intérêt scientifique est démontré, sur la base de leurs propriétés de pertinence, de validité et de fiabilité ;
- collecter et procéder au traitement des indicateurs, tout en mettant en place un accompagnement des établissements ;
- définir un cadre d'utilisation des indicateurs collectés, c'est-à-dire trouver une manière de formaliser les résultats, puis définir les modalités d'accompagnement des établissements dans leur utilisation à des fins d'amélioration de la qualité.

(5) Les Québécois ont traduit ce terme par « imputabilité », ce qui ne nous paraît pas totalement approprié dans le contexte français.

4.2. Un fort développement des incitations malgré un manque de stabilisation du cadre conceptuel

Des expériences étrangères émergent deux grands types d'incitations à la qualité : la diffusion publique des données de performances et les incitations financières. La question qui nous est aujourd'hui posée est de savoir s'il faut que le système français s'engage dans ces voies.

En ce qui concerne la diffusion publique d'informations, il ne s'agit de fait plus tant de savoir s'il faut ou non publier de l'information sur la qualité des établissements que de savoir comment bien le faire. Si les pouvoirs publics ne s'investissent pas dans l'information des usagers et des organismes de tutelle des établissements, d'autres le feront, sans nécessairement avoir le souci de la rigueur méthodologique et de la prudence à observer dans l'utilisation « externe » des indicateurs. Le premier constat, aujourd'hui unanimement admis, mais qui a largement surpris, est que la diffusion publique de l'information avait finalement peu d'impact direct sur le comportement des usagers. Ces derniers étaient pourtant la cible principale du système qui, dans une logique marchande, cherchait à réduire les asymétries d'information pour permettre aux consommateurs de sélectionner les producteurs offrant le meilleur rapport qualité/prix. Ce constat a provoqué un débat dans la littérature autour des explications de ce phénomène. Pour certains, la relative inefficacité pourrait tenir à l'inadéquation du contenu et des modalités de diffusion. Pour d'autres, tout effort supplémentaire pour actionner ce levier de la demande est voué à l'échec. Pour diverses raisons, le consommateur ne pourrait faire jouer sa rationalité : l'information fiable s'avérerait de toute manière toujours trop complexe, les usagers doivent en général effectuer un choix dans une situation plus ou moins urgente, leur regard porte plus sur le médecin que sur l'hôpital, etc. En revanche, l'impact sur les professionnels de santé est plus visible, montrant que c'est essentiellement un effet de réputation qui joue dans le cadre de la diffusion publique. Ces résultats engagent les pouvoirs publics à accompagner les professionnels pour qu'ils puissent s'approprier ces nouvelles logiques de comparaison, notamment en assurant leur cohérence avec les démarches plus anciennes d'amélioration continue de la qualité.

On sait en revanche encore peu de choses sur l'impact des incitations financières. L'absence d'éléments de connaissance ne permet pas de mettre en œuvre, à coup sûr, un système efficace tout en limitant les effets pervers. Malgré la diversité des expérimentations en cours outre-Atlantique, il n'est pas encore possible de déterminer *a priori* les « bons » éléments constitutifs de tels systèmes et beaucoup de questions restent encore sans réponse définitive : quel est le bon niveau de l'incitation, le professionnel ou la structure ? Comment doser l'incitation (quelle rémunération supplémentaire ou quelle diminution éventuelle) ? Faut-il rémunérer le niveau atteint ou l'évolution ? etc.

Beaucoup de questions restent encore ouvertes, mais il serait tout à fait regrettable de ne pas explorer cette voie, même s'il convient de procéder avec prudence en se laissant le temps de la recherche et de l'expérimentation.

Complément B

L'évaluation de l'enseignement supérieur et la recherche : un éclairage à partir du cas d'école de la réforme des études de licence, maîtrise et doctorat

Jean-Luc Gaffard

*Université de Nice Sophia Antipolis,
Institut universitaire de France et OFCE*

Les dysfonctionnements dont est l'objet le système français d'enseignement supérieur et de recherche sont, en général, bien identifiés. Études et rapports vont souvent dans le même sens s'agissant du diagnostic et des remèdes. En particulier, les dualités du système (entre écoles et universités, entre filières courtes sélectives et filières longues non sélectives) empêchent les meilleurs étudiants de suivre une formation par la recherche et conduit les moins aptes à tenter des filières longues pour lesquelles ils ne sont pas préparés. Les ressources sont particulièrement mal affectées. Le problème vient de ce que ces dysfonctionnements traduisent ce que les économistes pourraient appeler une coordination sur un mauvais équilibre institutionnel entre les bénéficiaires de différentes rentes. C'est bien cet équilibre qui rend le changement difficile à mettre en œuvre. Il est symptomatique de constater que la réforme engagée, suite, aux engagements européens correspondant au processus dit de Bologne, respecte le cadre institutionnel existant et ne concerne que la structure des études universitaires. Alors que le processus en question impliquait d'avancer, certes sur l'uniformisation des cursus, mais aussi et avant tout en développant (ou en instaurant) l'autonomie des universités et en promouvant des formes adaptées d'assurance qualité (en fait d'évaluation).

Les défaillances des modes d'évaluation propres au système institutionnel en vigueur expliquent pourquoi la réforme des études universitaires, loin de satisfaire les objectifs affichés, ne fait que souligner ces défaillances et n'apporte aucune solution au problème posé.

1. Les défaillances des modes d'évaluation

L'évaluation de la recherche et de l'enseignement supérieur est centralisée. Cette centralisation tient son unique raison d'être d'une organisation universitaire qui fait que les universités ne sont pas réellement responsables de leurs actes, ni en matière de politique scientifique ou pédagogique, ni en matière de recrutement : elles ne sont jamais sanctionnées si elles commettent des erreurs. Cette situation, ancrée dans les dispositifs institutionnels est la conséquence du choix de conduire une politique de scolarisation de masse dans l'enseignement supérieur dans le cadre d'établissements indistincts principalement organisés en cycles longs non sélectifs. Ce choix n'est en rien contraint comme le montre l'exemple de la Californie où le système universitaire public, diversifié et sélectif, conduit à un taux de scolarisation dans l'enseignement supérieur et un taux de délivrance des diplômes plus élevés, créant ainsi les conditions d'une croissance plus forte^(*).

L'évaluation de la recherche et de l'enseignement supérieur est centralisée, mais démultipliée entre organismes aux configurations et aux attributions distinctes : comité national d'évaluation, comité d'évaluation de la recherche, mission scientifique, technique et pédagogique, comité national de la recherche scientifique, comités de programmes de différentes natures et désormais l'Agence nationale de la recherche (ANR). Le propre de ces organismes est d'effectuer des expertises chacun de leur côté, sur des domaines similaires ou complémentaires, sans communication institutionnalisée entre eux, avec des moyens différents. Ce système repose sur les travaux d'experts dont les conditions de désignation et de travail (isolé ou en groupe) sont variables quant aux procédures et aux critères. Il établit une discrimination entre les équipes de recherche qui sont évaluées dans des conditions et suivant des critères différents suivant leur statut qui commande l'instance d'évaluation compétente. Il peut conduire à des écarts sinon à des contradictions de jugement et finalement à une confusion propre à rendre l'évaluation inopérante quant à ses effets.

L'évaluation de la recherche et de l'enseignement supérieur est conçue comme un service de l'administration centrale qui dispose du pouvoir de décision. C'est notamment le cas pour la Mission scientifique technique et pédagogique (MSTP). Ce qui est en cause, en l'occurrence, ce n'est pas la séparation de l'évaluation de la décision qui est nécessaire, mais l'opacité de l'évaluation pour les laboratoires ou les unités de formation évalués. Ainsi les avis circonstanciés de la MSTP sur les filières de formation ne sont pas transmis aux intéressés. Ils sont filtrés par la direction de l'enseignement supérieur et ses chargés de mission dont l'action duplique largement celle des experts de la MSTP et ne s'y conforme pas toujours. Les avis sur les laboratoires sont parfois filtrés par les présidents d'université. Les avis sur les dossiers déposés dans le cadre des programmes incitatifs sont filtrés par

(*) Cf. Belloc Bernard (2005) : « Le système public d'enseignement supérieur californien : un véritable service public ? Quels enseignements pour la France ? », *Fondation pour l'innovation politique*.

des comités *ad hoc*, sans même un retour de résultat. Il arrive certes que la MSTP parvienne à obtenir qu'il soit effectivement tenu compte de son travail d'expertise et d'évaluation, mais cela résulte avant tout de l'opiniâtreté et de l'autorité scientifique de ses membres. D'une manière générale, la multiplicité des échelons et l'opacité qui en résulte ne peuvent que favoriser des comportements opportunistes jouant sur les défauts d'information.

La conséquence de cette organisation de l'évaluation est triple. Il existe une forte inertie des décisions qui tendent à reproduire l'existant en termes de reconnaissance et de financement modulo des changements ponctuels qui peuvent être le fruit d'interventions parallèles (en dehors des règles) auprès des instances d'évaluation ou de décision correspondant à des réseaux politiques, corporatistes, voire simplement personnels. Il arrive, en effet, que des écarts de dotation injustifiés apparaissent qui traduisent ce type d'interventions. Il est certes possible de les corriger, mais ce ne peut être que de manière brutale et souvent conflictuelle compte tenu du rythme quadriennal des contrats. L'inertie qui prévaut généralement conforte la formation et la pérennité de rentes de situation, petites ou grandes, dont bénéficient des laboratoires ou des unités de formation dont le résultat est une allocation inefficace des ressources financières et humaines. Finalement, l'évaluation est sans conséquences, positives ou négatives, pour les évalués.

Le paradoxe du système héritier de situations de petit nombre est qu'il aboutit au renforcement de pouvoirs locaux (au sens large) qui, le plus souvent, n'ont pas de réelle légitimité ou assise scientifique dans le contexte institutionnel actuel (conditions d'élection ou de nomination des titulaires de ces pouvoirs) et qui peuvent agir au bénéfice de lobbys locaux plutôt que de servir une politique scientifique et pédagogique cohérente. En d'autres termes, ce système aboutit à l'inverse de ce pour quoi il a été fondamentalement conçu.

En bref, le mode actuel d'évaluation est un cas d'école des effets pervers des phénomènes de capture ou d'asymétrie d'information (aléa moral et anti-sélection) à tous les niveaux dont la conséquence ultime est une mauvaise allocation des ressources que ne peut évidemment pas corriger une augmentation de leur montant global.

Le pivot de la réforme nécessaire est, sans aucun doute, le mode de gouvernance des universités. Certes, il est toujours possible de rationaliser le système existant. La création de l'Agence nationale de la recherche ou la proposition d'un office national d'évaluation regroupant trois comités aux attributions complémentaires bien définies (évaluation des établissements, évaluation des personnels, évaluation des équipes, des unités de formation, des programmes et des relations internationales) sont censées aller dans ce sens. Encore faudrait-il que le changement institutionnel soit l'occasion de poser et de résoudre le vrai problème : celui de l'indépendance de l'évaluation fondée sur la transparence et la publicité des procédures et de l'information dont on voit mal comment il pourrait être dissocié du problème plus général de structure et de gouvernance de l'ensemble du système universitaire.

2. La réforme des études face aux difficultés de gouvernance du système universitaire

La réforme des études supérieures concernant l'ensemble des pays de l'Union européenne et fondée sur triptyque licence-maîtrise-doctorat (LMD) aurait dû être un moment privilégié de réflexion sur le futur de l'organisation universitaire, mais aussi sur celui des disciplines scientifiques, s'agissant à la fois de leur implication dans les dispositifs de formation et de leur capacité de développement au sein des universités. L'enjeu était et reste loin d'être mineur quand on sait que la réforme mise en œuvre ne peut qu'entraîner une transformation profonde des parcours de formation et une redéfinition de la carte universitaire pour de longues années. Les difficultés auxquelles se heurte, inévitablement, cette réforme tiennent aux comportements et aux modes de gouvernance.

Face aux enjeux de cette réforme, il aurait fallu pouvoir mettre un terme à un mode de gouvernance et des comportements qui favorisent la défense de rentes au détriment de l'intérêt collectif. Il aurait fallu que le cloisonnement des disciplines, qui peut s'expliquer par les conditions de leur constitution et les exigences de leur développement, ne conduise pas à un cloisonnement des formations dans le seul intérêt immédiat de leurs promoteurs. Il aurait fallu éviter que coexistent des formations avec de gros effectifs, mais dépourvus d'enseignements liés à la recherche, avec des formations à effectifs réduits organisées autour des thèmes de recherche en faveur dans la littérature académique. Il aurait fallu que l'adossement scientifique des formations de masters soit réel, impliquant une cohérence effective entre l'évaluation des équipes et l'habilitation des diplômés. Il aurait fallu, enfin, que la fonction universitaire puisse s'exercer pleinement au lieu de se réduire à la seule course aux colloques et aux publications, et donne lieu à des rémunérations correspondant aux compétences affichées.

Ces exigences n'avaient de chances d'aboutir que si des réformes avaient été conduites dans un environnement institutionnel dans lequel contraintes et incitations auraient conduit à prendre des décisions conformes à la fois à la qualité scientifique et aux besoins réels de formation.

Dans le contexte actuel, les résultats immédiats et prévisibles de la réforme ne pouvaient être qu'une illustration des défaillances du système d'évaluation.

Le choix de ne pas imposer de cadrage national des masters, revendiqué comme un gage d'ouverture de l'administration centrale, a conduit à une diversification incontrôlée des appellations de diplômes de même niveau et de même nature rendant la carte peu lisible pour les étudiants potentiels. Une telle diversification est le fruit d'un compromis par essence bureaucratique entre les présidents des universités et les bureaux de la direction des enseignements supérieurs laquelle ne dispose, par définition, que d'une information incomplète sinon biaisée et en tout cas non expertisée dans des

conditions de transparence et d'indépendance. Le cadrage qui s'est imposé a davantage respecté les intérêts corporatistes des groupes locaux dominant les conseils que les positions défendues par ceux qui, loin de ces conseils, s'attachent à développer des projets pédagogiques et scientifiques basés sur la réalité des disciplines et les besoins testés par l'expérience.

La dissociation entre l'évaluation des équipes et celle des masters a conduit la direction de l'enseignement supérieur et les universités à arrêter la carte des masters sur la base d'un compromis institutionnel qui témoigne du divorce entre les intérêts de la recherche et ceux de l'enseignement. Les contraintes statutaires de reconnaissance imposées par tel ou tel organisme de tutelle en recherche ou des majorités de rencontre des commissions de recrutement ont conduit à un équilibre non coopératif qui a pu obérer, soit le développement de thèmes de recherche déjà ancrés et intéressants des parcours de formation fortement demandés, soit, de manière générale, tout développement efficace de la recherche. Les universités ont souvent obtenu des habilitations de masters, soit adossés à des recherches au périmètre restreint impliquant des effectifs réduits, soit sans adossement scientifique éprouvé. Elles se sont ainsi placées dans la situation inconfortable et surtout sans rapport avec les exigences de service public de proposer une offre tronquée ou de second ordre.

De manière générale, dans l'environnement institutionnel actuel, nombreux ont été les avis de la MSTP qui n'ont pas été suivis ou ont été reformulés par la direction de l'enseignement supérieur comme conséquence de compromis passés directement avec les établissements, ou dans quelques cas avec les responsables de disciplines au niveau national. Le résultat le plus tangible a été une augmentation sensible des deuxièmes années de master vis-à-vis des anciens DEA et DESS (+ 30 % en moyenne) impliquant de conduire un plus grand nombre d'étudiants à bac + 5 sans réelle sélectivité. Certaines des avancées réalisées avec la mise en place des DESS ont ainsi été perdues avec la duplication des filières dans une même université ou dans des universités voisines. Des formations de master non adossées à la recherche ont été créées dans des domaines où le niveau licence aurait dû constituer le niveau terminal. Certes, les instances d'évaluation ont effectué un classement des formations et ainsi repéré celles qui n'auraient pas dû être habilitées. Il est, en outre, vraisemblable que la contrainte budgétaire va jouer et conduire à une réduction du nombre des masters après quelques années d'exercice. Mais cette contrainte opérant au niveau des établissements, la sélection des masters, au lieu d'être fondée sur les expertises indépendantes, traduira le jeu des influences locales et n'aboutira pas toujours à retenir ceux des masters qui sont véritablement adossés à une recherche de qualité et qui correspondent à des besoins réels de formation à ce niveau.

3. Pistes pour une réforme de la structure et de la gouvernance de l'enseignement supérieur

En bref, la réforme des études mise en œuvre a surtout révélé les dysfonctionnements existants faute de s'être appuyée sur un changement des modes de gouvernance de l'ensemble du système d'enseignement supérieur et de recherche.

Ce changement devrait avoir pour objectif de substituer au système actuel fondé sur l'opposition entre écoles et universités un système hiérarchisé, ouvert et sélectif. Il distinguerait trois types d'établissement : les établissements délivrant licence, maîtrise et doctorat et abritant les équipes de recherche (les universités de recherche), sans doute au nombre d'une vingtaine ; les établissements délivrant licence et maîtrise (les universités) ; les établissements délivrant seulement la licence (les collèges universitaires). Ces établissements, pleinement autonomes, seraient gérés par des conseils d'administration d'une quinzaine de membres, composés de représentants des personnels enseignants et chercheurs, de représentants de l'État et de la Région, et de personnalités scientifiques extérieures, et désignant le président. Le contrôle et l'évaluation seraient assurés pour chaque groupe d'établissement, au niveau géographique qui convient (national ou régional), sous la responsabilité d'un conseil composé de personnalités extérieures nommées et titulaires du diplôme le plus élevé délivré par les établissements concernés.

La phase de transition impliquerait que soient créées les conditions d'une restructuration des différents établissements actuellement engagés dans l'enseignement supérieur et la recherche : fusions d'écoles et d'universités, constitutions de collèges universitaires à partir des IUT et BTS, développement de classes préparatoires intégrées.

Le choix d'un mode de gouvernance de ce type, en créant les conditions d'une évaluation mieux ciblée, plus transparente, aurait pour premier avantage de réduire considérablement les phénomènes de rente et de favoriser une meilleure allocation des ressources au regard des objectifs fixés d'accroître le niveau de recherche et de développer les compétences favorables à la croissance.

Complément C

La mesure de la production et de la productivité des administrations publiques : les différentes approches, en particulier celle de la mission Atkinson au Royaume-Uni

Sylvie Hel-Thelie

Direction générale du trésor et de la politique économique

La mesure de la production et de la productivité des administrations publiques a suscité différents travaux, depuis le début des années quatre-vingt. Les approches – et les objectifs – de ces travaux sont toutefois différentes. Les travaux les plus récents sont ceux de la mission Atkinson, au Royaume-Uni, dont le rapport a été publié le 31 janvier 2005.

Cette note rappelle le contexte actuel ainsi que les différentes approches de la mesure de la production et de la productivité des administrations publiques, avant de s'intéresser plus particulièrement aux résultats de la mission Atkinson.

1. Les travaux sur la mesure de la production et de la productivité des administrations publiques répondent à des préoccupations diverses

1.1. Des réflexions déjà anciennes

Sylvie Trosa (2004) rappelle que certains pays (Nouvelle-Zélande, États-Unis, mais aussi Royaume-Uni) ont développé une approche du management du secteur public par *outputs* (= production de services, prestations immédiates) à partir du début des années quatre-vingt. Ces pays ont depuis renoncé à un management exclusivement par *outputs* et gèrent aujourd'hui à la fois par *outputs et outcomes*, c'est-à-dire en considérant aussi les

effets recherchés. Si la première approche n'a pas eu que des effets négatifs (intérêt porté à la qualité des prestations offertes, mise en place d'outils de comptabilité et de systèmes d'information performants), elle a aussi été critiquée pour la démotivation des fonctionnaires qu'elle induisait, l'important étant de mesurer plus que de comprendre le sens de la mesure. Juger les agents uniquement sur les *outputs* les incite à ne faire que ce qui leur est demandé, sans effort supplémentaire.

En France, l'approche en la matière était peu développée avant la mise en place de la LOLF. On peut cependant mentionner les travaux qui avaient été développés par le CERC (1994), dans la seconde moitié des années quatre-vingt, avec l'idée de mettre en place un outil de gestion prévisionnelle dans les services publics, sur la base de la mesure de la production (les services rendus), de la performance productive globale et du partage des éventuels gains de productivité entre les partenaires (usagers, salariés, fournisseurs, tutelle...). Cet outil n'avait pas pour objectif de tester la productivité d'une politique (éducation, santé...) dans son ensemble mais celle des entités publiques qui les mettaient en œuvre. Il a ainsi été mis en place et apparemment utilisé dans certains services publics : CNAV, CAF, CNOUS et CROUS, services techniques des directions départementales de l'Équipement, etc. Ce type d'outil et de mesure n'a toutefois jamais été généralisé au sein de l'administration française.

L'une des ambitions de la LOLF est de permettre la mesure des *outcomes* (impact des politiques/objectifs) comme de l'efficacité (c'est-à-dire des résultats par rapport aux moyens mis en œuvre), mais sous une forme non systématique, tout au moins s'agissant de la partie émergée de la production d'indicateurs (indicateurs de performance figurant dans les documents budgétaires) et uniquement pour l'État. Le pilotage des programmes instaurés par la LOLF devrait cependant susciter également la mise en place d'une panoplie de mesures et d'indicateurs de gestion qui devrait, entre autres, permettre d'approcher une mesure de la productivité des services produits par les administrations.

Enfin, de manière beaucoup plus agrégée, certains travaux récents cherchent à apprécier les marges de manœuvre financières dont peuvent disposer les pouvoirs publics des pays grâce à la réalisation de gains de productivité. Ces approches comparent l'efficacité des administrations de différents pays en reliant la production/la performance des administrations publiques aux dépenses publiques mises en œuvre. Les méthodes utilisées permettent d'estimer un « taux d'efficacité des dépenses » correspondant à la part de dépenses en moins dont un pays aurait besoin pour réaliser le même niveau de production/de performance. Pour l'Union européenne à 15, le taux moyen est estimé à 0,73 pour les années quatre-vingt-dix, ce qui signifie que le même niveau de production pourrait être atteint en utilisant seulement 73 % des ressources qui sont utilisées actuellement (i.e. en économisant 27 % de dépenses publiques). Pour la France, ce taux est estimé à 0,64 (soit une économie potentielle de 36 % des dépenses publiques) (Afonso,

Schuknecht, Tanzi, 2003). Les résultats de tels travaux doivent cependant être pris avec précautions, d'abord parce qu'ils comportent des approximations⁽¹⁾, notamment en distinguant mal ce qui relève de la production (*outputs*, résultats immédiats) de la performance (*outcomes*, impacts indirects des politiques) dans les données qu'ils utilisent, ensuite parce qu'ils sont très agrégés et ne fournissent évidemment pas de précisions sur les secteurs les plus concernés ou la manière de concrétiser ces gains de productivité⁽²⁾.

Les travaux d'amélioration de la mesure de la production et de la productivité des administrations publiques, notamment ceux engagés par la mission Atkinson, poursuivent un objectif plus technique en apparence – permettre une meilleure appréciation de la production de services publics dans les comptes nationaux – mais constitue en fait une voie parmi d'autres pour répondre aux interrogations des citoyens et aux préoccupations du gouvernement sur l'utilisation des ressources publiques.

1.2. La Commission européenne cherche à améliorer l'analyse économique et budgétaire réalisée par les États membres

La valeur de la production – essentiellement non marchande – des administrations publiques est, jusqu'à présent, évaluée par les comptes nationaux de manière conventionnelle par son coût. Il n'existe en effet pas de prix de marché pour des services publics tels que la police, l'enseignement... Il n'est donc pas possible de mesurer la productivité d'un service public (mesurée comme la relation entre les moyens mis en œuvre et la production réalisée) puisque, par définition, la valeur de la production non marchande de biens et de services des administrations publiques est estimée égale aux dépenses réalisées pour les produire.

Mieux évaluer la valeur de la production non marchande de l'État représente pourtant un enjeu important dans la mesure où les services publics (ou plutôt le coût de ceux-ci mesuré par le PIB non marchand) peuvent représenter 20 à 25 % du PIB.

Les services de la Commission européenne cherchent à améliorer la prise en compte de la production de services publics dans les comptes nationaux. La Commission a ainsi demandé aux États membres, en décem-

(1) Les auteurs soulignent eux-mêmes que les résultats dépendent de la sélection des indicateurs mesurant la production (ou la performance : les données relèvent selon les cas de l'une ou de l'autre acception), de la non pondération de ces indicateurs pour mesurer la performance globale des administrations publiques d'un pays, du choix de rapporter les dépenses au PIB pour mesurer les moyens mis en œuvre.

(2) Des résultats ont cependant également été obtenus par secteurs (*cf.* Afonso et St. Aubyn, 2004). Dans le cas de l'éducation, par exemple, les résultats obtenus montreraient que la même production pourrait être réalisée en utilisant, selon les pays, entre 11 et 48 % de dépenses en moins qu'actuellement. La production est, dans ce cas, approchée par les résultats des études PISA, qui sont plus des indicateurs de performance que des indicateurs de production (au sens de résultats immédiats par rapport à impact d'une politique).

bre 2002⁽³⁾, que les comptes nationaux soient établis à partir de 2006 en estimant les services publics *individuels* à leur valeur d'usage et non plus en fonction des moyens mis en œuvre. Dans le domaine de la santé, cela reviendrait par exemple à prendre en compte le nombre de patients traités au lieu de considérer le nombre de médecins impliqués. Les changements en termes de qualité (dans cet exemple, la qualité des traitements médicaux) devraient également être reflétés dans la mesure de la production.

L'objectif poursuivi par la Commission européenne (ainsi que par les autres institutions internationales qui ont édicté des règles dans ce domaine) est de pouvoir mieux analyser la situation économique et budgétaire d'un pays et de pouvoir la comparer à celle d'autres pays en disposant de comptes nationaux élaborés sur la base de principes uniques et non diversement interprétables.

La France devrait commencer à adopter cette approche dans la nouvelle base 2000 qui sera mise en œuvre pour la réalisation du compte provisoire des administrations publiques de l'année 2004 (disponible en juin 2005), tout au moins dans un premier temps pour les secteurs de l'éducation et de la santé (cf. encadré 1). Parmi les États européens, il semble toutefois que ce soit le Royaume-Uni qui ait abordé cette question de la manière la plus approfondie.

1. La mesure de la production non marchande dans la base 2000

Dans la nouvelle base, la production non marchande sera toujours mesurée par le PIB non marchand, donc en valeur. Le changement par rapport à l'ancienne base résidera dans le partage volume/prix réalisé pour certains services individuels. Jusqu'à présent, pour la mesure de la production en matière d'éducation, l'indice de prix était approximé par l'évolution du point de la Fonction publique et le volume en était déduit automatiquement. Dans la nouvelle base, et pour les dépenses d'éducation, l'évolution du volume de production non marchande d'éducation sera directement estimée sur la base du nombre d'élèves accueillis (par niveau : primaire, secondaire, supérieur) pondéré par les coûts relatifs.

Ce même calcul direct devrait également être réalisé pour le secteur de la santé. Il semble cependant difficile d'étendre pour le moment cette pratique à d'autres services publics, leur production étant considérée comme moins facilement directement mesurable par les statisticiens français.

(3) C(2002) 5054/F du 17/12/2002 – 2002/990/CE : décision de la Commission clarifiant davantage l'annexe A du règlement (CE) n° 2223/96 du Conseil en ce qui concerne les principes de la mesure des prix et des volumes dans les comptes nationaux.

2. Les travaux de la mission Atkinson s'attachent à améliorer la mesure de la production et de la productivité des administrations publiques britanniques sur la base de la comptabilité nationale

2.1. Les travaux britanniques sur la mesure de la productivité directe sont déjà bien engagés

L'Office national de statistiques britannique (ONS) a demandé en décembre 2003 à un groupe de travail indépendant créé à cette occasion, la mission Atkinson, de formuler des règles précises d'évaluation de la production, pour chacun des grands services publics. Cette mission a remis un rapport intérimaire en juillet 2004 et a fourni son rapport définitif le 31 janvier 2005.

Les travaux de la mission se sont appuyés sur ceux de l'Office national de statistiques britannique (ONS) qui a commencé, depuis 1998 à renouveler la mesure de la production des administrations publiques au Royaume-Uni. De fait, les estimations directes de la production représentent actuellement les deux tiers de la consommation finale des administrations publiques au Royaume-Uni et sont publiées annuellement.

L'approche mise en œuvre par les comptables nationaux britanniques, s'appuyant en cela sur les recommandations de la Commission européenne et les règles de bonnes pratiques établies au niveau international, est de mesurer le volume de la production non marchande des administrations publiques de manière directe, indépendamment des dépenses correspondantes. Ainsi, le volume de production du secteur de la santé sera mesuré en termes de nombre de traitements délivrés aux patients ; le volume de la production du secteur de l'éducation sera mesuré par le nombre d'heures de cours prodiguées aux élèves.

Concrètement, la mesure directe de la production est réalisée au niveau sectoriel⁽⁴⁾ à l'aide des indicateurs définis (cf. tableau 1). À partir d'une année (et d'un niveau de production) de référence (en l'occurrence, l'année 1995), on estime l'évolution de la production comme la croissance (ou la décroissance) de la mesure directe réalisée par les indicateurs. L'agrégation des mesures sectorielles de la production est ensuite réalisée en pondérant les mesures de volume par les coûts relatifs de production de chacun des secteurs.

La mesure de la productivité des administrations publiques découle de cette nouvelle approche de la mesure de la production : elle résulte du rapport entre le volume de production et le volume de moyens utilisés pour réaliser cette production, pour une période donnée.

(4) Santé, éducation, protection sociale (sécurité sociale et services à la personne), ordre public et sécurité (police, justice, sécurité civile, secteur pénitentiaire). Ces secteurs font partie des dix grandes fonctions de la classification COFOG (nomenclature des Nations unies de classification des dépenses publiques).

Grandes fonctions de dépenses et exemple d'indicateurs associés

Fonctions ^(*)	Dépenses des admin. publiques britanniques (en %, année 2000)	Exemples d'indicateurs de production retenus
Santé	30,3	Nombre de consultations médicales, index d'activité hospitalière
Éducation	17,1	Nombre d'élèves dans le primaire, le secondaire, ajustement prenant en compte la qualité
Protection sociale <ul style="list-style-type: none"> • Sécurité sociale • services sociaux à la personne 	2,7	Nombre d'allocataires pour les douze principales allocations
	7,4	Nombre d'enfants et d'adultes gérés
Ordre public et sécurité <ul style="list-style-type: none"> • Police • Justice • liberté surveillée • prisons • sécurité civile 	5,8	Volume d'activité policière (nombre de crimes, nombre d'accidents, nombre de patrouilles, etc.)
	3,0	Nombre d'affaires devant les tribunaux pondéré par le nombre d'heures ou les coûts moyens
		Heures de travail des professionnels concernés
		Nombre de détenus
1,1	Nombre d'incendies traités, prévention des incendies	

Note : Pour les autres fonctions, le volume de la production reste par définition égal aux dépenses correspondantes (rémunérations, consommation en capital,...) déflatées.

Source : Rapport Atkinson.

À l'heure actuelle, les estimations directes agrégées de production des administrations publiques sont officiellement publiées par l'ONS (Office national de statistiques britannique), bien que toujours soumises à des travaux d'amélioration, portant notamment sur la prise en compte des modifications de la qualité de la production. Elles représentent déjà actuellement les deux tiers de la consommation finale des administrations publiques au Royaume-Uni.

En revanche, les travaux portant sur la mesure du volume des moyens mis en œuvre ainsi que sur la mesure de la productivité qui en découle sont à ce stade encore expérimentaux. Ils sont néanmoins publiés depuis 2002 dans la revue de l'ONS (*Economic Trends*). Contrairement au rapport intérimaire présenté en juillet 2003, le rapport final ne reprend pas ces éléments. Selon les travaux de l'ONS prenant l'année 1995 comme base, la productivité des administrations publiques au Royaume-Uni aurait augmenté de 1,5 % jusqu'en 1998 puis chuté de près de 4 % entre 1998 et 2001 (*cf.* annexe 1).

2.2. Le rapport Atkinson pose des principes et étudie leur mise en œuvre dans divers secteurs

Dans ce contexte, l'objectif du rapport de la mission Atkinson est en premier lieu de formuler des principes (9) et des recommandations méthodologiques (54 au total) relatives à la mesure des *inputs* (moyens engagés), *outputs* (production) et productivité dans les comptes nationaux (voir encadré 2 précisant ces différentes notions sur la base d'un exemple). À cette fin, le rapport examine successivement différents aspects en relation avec la question posée⁽⁵⁾ : mesure actuelle de la production des administrations, contexte international des travaux, fonctions et limites des comptes nationaux (chapitres 2 et 3). Le chapitre 4 constitue le cœur du rapport puisqu'il traite du cadre méthodologique de la mesure directe de la production ainsi que de la productivité et propose neuf grands principes pour son amélioration. Les chapitres 5 et 6 s'intéressent au processus de mise en œuvre pratique des principes posés dans l'administration statistique britannique, respectivement pour les *inputs* (dépenses) et les *outputs* (production). Le chapitre 7 examine quelques problèmes généraux relatifs aux changements à mettre en œuvre.

La mission Atkinson a également pour objectif de traduire ces principes et recommandations méthodologiques au niveau sectoriel. Plusieurs chapitres (8 à 11) passent en revue les principaux secteurs de dépenses des administrations publiques (santé, éducation, ordre public et sécurité, protection sociale) pour la mise en œuvre de la méthodologie proposée.

2. Les notions d'*input*, *output*, *outcome* distinguées dans le Handbook d'Eurostat

Les mesures de la productivité mettent en relation *outputs* et *inputs*. Le problème principal est celui de la mesure directe de la valeur de marché de la production (*output*), les services publics n'ayant pas de vrai prix de marché qui refléterait la valeur des services, ceux-ci étant fournis aux usagers gratuitement ou à un prix subventionné

En prenant le secteur de la santé comme exemple, on identifie les *inputs* (ou moyens engagés, dépenses) comme le temps passé par l'équipe médicale et non médicale, les médicaments, l'électricité et les autres dépenses réalisées, ainsi que les services en capitaux fournis par les équipements et les bâtiments utilisés. Ces ressources sont utilisées dans l'exercice de la médecine comme dans les activités hospitalières. Elles fournissent du service « santé publique » bénéficiant aux patients de manière individuelle. Le service « santé » constitue l'*output* (la production directe) associé aux activités d'*input*. L'*outcome* « santé » (l'impact, la production indirecte), correspondant à l'état de santé de la population concernée, dépendra en revanche de facteurs externes à la production du service de santé, tel que l'usage du tabac, par exemple.

(5) Cf. annexe 2 pour un résumé du contenu des premiers chapitres du rapport.

2.3. Un travail qui devrait impulser des avancées méthodologiques utiles

Le rapport Atkinson constitue sans aucun doute un ensemble extrêmement utile pour les statisticiens des autres États membres (et des autres pays développés) par les différents aspects qu'il aborde et par les recommandations – générales ou spécifiques – qu'il formule.

Les travaux menés dans le cadre de ce rapport au sein des administrations britanniques concernées vont permettre des avancées bienvenues dans la mesure directe de la production des administrations. C'est en particulier le cas d'une meilleure prise en compte de la qualité dans la mesure de la production de certains secteurs (éducation, santé, notamment). Dans certains cas, les travaux devraient conduire à utiliser une méthode sensiblement différente de la méthode actuelle (reposant déjà sur une mesure directe de la production). Dans d'autres cas, il s'agira plutôt d'améliorer la méthode existante.

Le rapport souligne l'importance à accorder à la présentation des mesures de la production des administrations et de leur productivité dans les comptes nationaux, en faisant ressortir les intentions pour lesquelles elles peuvent – ou ne peuvent pas – être utilisées. Ces mesures sont différentes des mesures microéconomiques de l'efficacité du secteur public, même si elles devraient être cohérentes avec celles-ci.

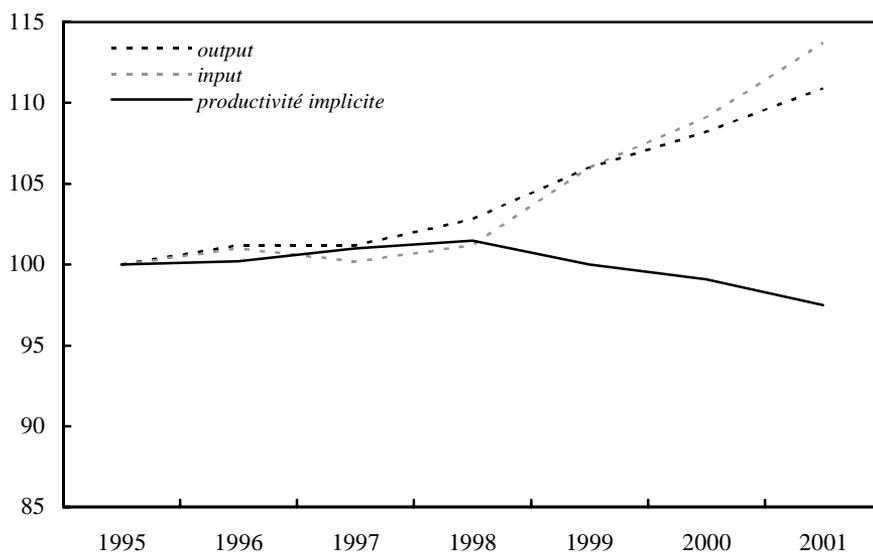
Au total, le principal intérêt de ce travail, portant sur les principes et les méthodes, sera sans doute la forte impulsion que cette réflexion aura donnée à l'amélioration des mesures dans les différents secteurs concernés de la production des services publics. La mission Atkinson, appuyée par une équipe de statisticiens localisée à l'ONS et l'ONS elle-même, a en effet travaillé en relation avec les principaux ministères responsables de la mesure des services publics. Ces travaux ont permis d'identifier un certain nombre d'avancées potentielles relatives à la manière de mesurer la production du secteur public dans le cadre des comptes nationaux. En ce sens, on peut penser que le rapport de la mission Atkinson, centré autant sur le « comment faire ? » que sur l'établissement de principes, aura véritablement été utile au gouvernement et à l'administration britanniques.

Annexe 1

Estimations récentes de la production directe des administrations et de leur productivité implicite

En millions de livres sterling

	1995	1996	1997	1998	1999	2000	2001
Volume de la production des administrations ^(*)	141,031	142,702	142,779	144,991	149,419	152,524	156,361
Volume de la production des administrations ^(**)	100,0	101,2	101,2	102,8	106,0	108,2	110,9
Volume des dépenses des administrations ^(*)	141,031	142,388	141,371	142,785	149,441	153,877	160,320
Volume des dépenses des administrations ^(**)	100,0	101,0	100,2	101,2	106,0	109,1	113,7
Index implicite de productivité ^(**)	100,0	100,2	101,0	101,5	100,0	99,1	97,5



Notes : (*) Aux prix 1995 ; (**) Index 1995 = 100.

Source : *Economic Trends*, juillet 2003, tableau 2.

Annexe 2

Résumé des chapitres généraux (1 à 7) du rapport de la mission Atkinson

1. Le chapitre 1

Il est consacré à une présentation générale du contenu du rapport. Il traite aussi du rôle des comptes nationaux du point de vue économique ainsi que de leurs limites. Les principales conclusions de ce chapitre sont les suivantes :

- le rapport traite de méthodologie et ne présente pas de nouvelles données relatives à la production ou la productivité des dépenses publiques. Son objectif est d'établir la future stratégie de travail dans ce domaine ;
- la construction des mesures de production des administrations publiques dans le cadre de la comptabilité nationale dépend beaucoup de la coopération des ministères concernés ;
- les comptes nationaux ont différents objectifs, et un seul résultat ne peut servir à répondre à tous. Différents agrégats permettront de répondre aux différentes questions ;
- le revenu national est un indicateur de la contribution au bien-être (*welfare*) d'activités économiques spécifiques. Ce n'est pas une mesure du bien-être économique total. Le bien-être n'est pas le seul objectif des politiques publiques ;
- les estimations des comptes nationaux pour les administrations publiques sont liées, mais différentes, des mesures microéconomiques de performance des administrations publiques, et ont des objectifs différents. Vouloir les utiliser de la même façon que des cibles de performance du secteur public serait méconnaître leur nature et leurs limites. Les mesures de performance exigent des indicateurs précis, transparents et simples à mesurer, non sujets à manipulation, mais ils ne doivent pas forcément être stables au cours du temps et peuvent être sélectifs dans leur couverture. À l'inverse, les mesures de production pour les comptes nationaux nécessitent d'être aussi complètes que possible et stables au cours du temps. C'est ce qui distingue les *outputs* des *outcomes*. Les comptes nationaux fournissent des indicateurs des grandes tendances. Les indicateurs issus des comptes nationaux doivent être cohérents avec les résultats des études de performance, mais ils n'en sont pas un substitut ;

- les comptes nationaux sont construits sur une série de conventions qui sont soumises à des marges d'erreur variant selon les différents domaines des comptes nationaux. L'importance donnée à ces erreurs dépend de l'objectif pour lequel les données sont utilisées.

2. Le chapitre 2

Il traite de la mesure de la production des administrations au Royaume-Uni. À la différence du rapport intérimaire de juillet 2003, le rapport final ne donne pas de tableaux et de graphiques précis relatifs à l'évolution de la production et de la productivité implicite des administrations publiques (tels que présentés dans l'annexe 1), sans doute pour ne pas concentrer l'attention des lecteurs de ce rapport sur les chiffres mais au contraire pour mettre l'accent sur la méthodologie. Les conclusions de ce chapitre sont les suivantes.

La mesure de la productivité des administrations suppose de mesurer le volume (prix constants) de la production (non marchande) et de le relier au volume des moyens mis en œuvre (dépenses publiques). Dans la plupart des pays et jusqu'en 1998 pour le Royaume-Uni, la valeur de la production des administrations était, par convention, égale à la valeur des *inputs* (entrées ou dépenses, telles que les salaires des employés, le coût des biens et services, la prise en compte de la consommation de capital fixe), et le volume de la production était le volume des *inputs*. Cette convention négligeait de fait les modifications de productivité :

- à partir de 1998, l'ONS a introduit progressivement le remplacement de l'approche de mesure conventionnelle de la production ($output = input$) par des mesures directes du volume de la production des administrations publiques. Les estimations directes couvrent actuellement les deux tiers de la consommation finale des administrations publiques ;

- de l'expérience des premières tentatives dans les années cinquante et soixante, on peut conclure que la mise en place des mesures directes de production demande beaucoup d'attention et requiert des investissements importants (en termes de moyens humains). Les mesures directes de la production doivent en effet être régulièrement vérifiées pour s'assurer qu'elles prennent bien en compte les modifications de qualité. L'ONS doit être attentive aux changements de politiques et garantir l'indépendance de l'approche mesurant la production publique ;

- les changements institutionnels au sein des administrations publiques entraînent des difficultés pour la mesure de la production. Les changements technologiques sont difficilement pris en compte ;

- la mesure de la productivité implicite des administrations est obtenue à partir de trois éléments : les dépenses (ou *inputs*), l'index de prix des dépenses et la mesure de la production directe (*output*). La fiabilité de ces trois éléments nécessite d'être évaluée. L'augmentation de la productivité sera sous-estimée si la mesure des dépenses est trop élevée, ou si la mesure de la production ou de l'augmentation des prix est trop faible, et sera surestimée inversement. Une grande importance doit être attachée à la mesure

des *inputs*. Pour la mesure de la productivité, la prise en compte de la qualité est importante à la fois pour les *inputs* et pour les *outputs* ;

- le travail de rassemblement et d'agrégation des données nécessitant une collaboration entre l'ONS, le Treasury et les autres ministères, est complexe et demande des travaux approfondis.

Si le Gouvernement britannique souhaite disposer d'estimations fiables de la production et de la productivité des administrations publiques, les ressources en matière statistique doivent être mises à disposition, à la fois à l'ONS et dans les ministères.

3. Le chapitre 3

Il s'intéresse au contexte international. Il existe en effet un certain nombre de règles de bonnes pratiques (*United Nations System of National Accounts – SNA 1993* –, *OECD Productivity Manual*, *European System of Accounts 2001* – ESA 1995) sur la mesure directe de la production. Il ressort de cet examen que :

- le programme de travail de l'ONS depuis 1998 est en ligne avec les recommandations SNA 1993. Le SNA recommande de distinguer la production de services individuels (services de santé, par exemple) dont la mesure est plus facile, des services collectifs (comme la défense) dont la mesure est plus délicate. D'autres pays se sont engagés dans cette même démarche à un moindre degré que le Royaume-Uni. D'un examen mené par l'OCDE de l'état d'avancement des mesures de productivité directe au sein des pays appartenant à l'Organisation, il ressort que le Royaume-Uni est l'un des pays les plus avancés dans ce domaine (deux tiers des dépenses couvertes, priorité au développement de cette dimension des comptes nationaux).

- il serait souhaitable que le travail de l'ONS influence le SNA 2008.

- l'ESA 1995 ainsi que le *Eurostat Handbook on Prices and Volumes Measures in National Accounts* ont étendu les recommandations du SNA en introduisant une classification des méthodes en A (recommandées), B (acceptables faute de mieux) et C (à proscrire). Sous réserve que le volume de chaque indicateur soit estimé séparément, l'ONS pourrait continuer sans problème à utiliser la mesure des *inputs* pour les services collectifs. Pour les services individuels, en revanche, les mesures d'*inputs* sont considérées comme inacceptables. Conformément à la décision de la Commission européenne du 17 décembre 2002, à compter de 2006 (à l'exception du Danemark qui bénéficie d'une dérogation jusqu'en 2012), les méthodes de type C ne seront plus acceptables. Dans le cas des services individuels, cela exclut l'utilisation de la méthode *output = input* utilisée par convention pour mesurer la production des administrations. Dans le cas des services collectifs, les mesures d'*input* peuvent être conservées (considérées comme une méthode de type B) sous réserve qu'elles satisfassent à certaines conditions ;

- nombre de pays travaillent actuellement sur le développement de mesures directes de la production, incluant le traitement des modifications de qualité, et il serait souhaitable que des coopérations soient établies dans ce domaine.

4. Le chapitre 4

Il constitue le cœur du rapport car il construit un cadre méthodologique et fixe les neuf grands principes présentés dans la note. La mesure directe de la production à partir des dépenses des administrations – et la mesure des *inputs* et de la productivité – doit être fondée sur un ensemble de principes, à l'intérieur du cadre établi par les recommandations internationales. Ces neuf principes sont les suivants.

4.1. Principe A

La mesure de la productivité des administrations publiques devrait, autant que possible, suivre des lignes directrices analogues à celles adoptées dans les comptes nationaux pour la production du secteur marchand.

4.2. Principe B

La production des administrations publiques devrait, en principe, être mesurée de manière à intégrer les modifications de qualité, prenant en compte la contribution du service considéré à l'*outcome* (résultat indirect de la dépense). Dans l'exemple de l'éducation cela revient à mesurer l'amélioration des résultats individuels attribuables au service public d'éducation.

4.3. Principe C

Il convient de tenir compte de la complémentarité entre production publique et privée, permettant l'augmentation de la valeur réelle des services publics dans une économie où le PIB réel s'accroît.

4.4. Principe D

Des critères officiels doivent être mis en place pour l'extension des mesures de production directe à de nouveaux secteurs. Les conditions pour introduire un nouvel indicateur de mesure de la production directe doivent être :

- qu'il couvre de manière adéquate l'ensemble des services pour le secteur considéré ;
- qu'il tienne compte des modifications de qualité ;
- que les effets de son introduction aient été testés service par service ;
- que le contexte dans lequel il sera publié ait été bien expertisé, en particulier s'agissant de l'estimation de la productivité qui en découle ;
- qu'une révision statistique régulière soit prévue.

4.5. Principe E

Les mesures doivent couvrir l'ensemble du Royaume-Uni. Quand les systèmes de fourniture des services publics et/ou les données disponibles diffèrent entre les pays faisant partie du Royaume-Uni, il est nécessaire que le choix des indicateurs reflète cette diversité.

4.6. Principe F

La mesure des *inputs* devrait être aussi complète que possible et devrait, en particulier, inclure les services en capitaux. Les *inputs* liés au travail devraient être recensés en utilisant à la fois des méthodes directes et indirectes, comparées et réconciliées.

4.7. Principe G

Des critères doivent être établis concernant la qualité des déflateurs de prix à appliquer aux séries de dépenses (*inputs*). Ils doivent être suffisamment désagrégés pour prendre en compte les changements dans la composition des dépenses et devraient refléter les coûts réels et complets.

4.8. Principe H

Des éléments indépendants et confirmés devraient être recherchés sur la productivité des administrations, participant à un processus de test des résultats et prenant notamment en compte l'éventuel décalage entre les moyens (*inputs*) mis en œuvre et la production obtenue (*outputs*). (cf. encadré 3).

4.9. Principe I

Une référence explicite devrait être faite aux marges d'erreur affectant les estimations issues des comptes nationaux.

Application du principe H au secteur de la santé

L'estimation de la productivité doit – ainsi que le prévoit le principe G – être testée contre d'autres éléments. Cela peut notamment passer par :

- l'examen des tendances de la durée moyenne de séjour à l'hôpital : une baisse de cette durée pourrait indiquer une amélioration de la productivité ;
- l'observation des changements de qualifications dans les équipes sociales : une augmentation du niveau d'ensemble des qualifications laisse légitimement attendre une croissance de la qualité de la production ;
- l'examen de la migration des traitements d'un dispositif onéreux (hôpital spécialisé) vers un dispositif moins onéreux (dispensaire) ;
- l'examen des modifications dans la technologie pour traiter des affections particulières, en particulier dans les domaines où de tels développements font passer la production d'une catégorie de dépenses à une autre (substitution d'un traitement médicamenteux à la chirurgie, par exemple) ;
- l'examen du niveau d'utilisation des ressources rares, telles que les lits de soins intensifs, les ambulances ou les salles d'opération : un taux élevé d'utilisation peut être aussi inefficace qu'un taux très bas, si du temps est consommé pour trouver une solution alternative et/ou si les patients en subissent des effets négatifs.

5. Le chapitre 5

Il s'intéresse aux *inputs* et aux déflateurs. Différentes recommandations sont formulées sur la manière très concrète d'obtenir et de classer les données, appliquée au cas britannique. Une étude de cas spécifique relative au secteur de la santé est présentée à titre d'exemple. Un schéma précis de traitement des données est enfin décrit. Les problèmes détectés sont analysés et leur résolution recherchée.

6. Le chapitre 6

Il effectue le même travail pour les *outputs*, illustrant les problèmes rencontrés par des exemples choisis dans des secteurs de dépenses spécifiques et donnant des recommandations sur les différents points abordés.

Deux aspects doivent être considérés séparément : les changements qui doivent être réalisés dans les secteurs de dépenses où les mesures directes de la production existent déjà ; la question de savoir si ces mesures directes de la production doivent être étendues à de nouveaux secteurs de dépenses.

Dans le premier cas, des recommandations précises d'amélioration des mesures sont faites, à partir de l'existant. La classification des services publics entre collectifs et individuels doit être précisée. Il convient de réfléchir aux pondérations utilisées dans la construction d'un indicateur de production unique. Les différents items (mesurés par leur taux d'évolution à partir d'une année donnée) doivent être combinés (selon les règles fixées par le SNA) en les pondérant par leur importance économique telle que mesurée par leur valeur. Les coûts disponibles sont cependant souvent des coûts moyens alors qu'il faudrait disposer de coûts marginaux.

D'autres problèmes de mesure peuvent se poser quand des services publics génèrent plus d'un « produit », ou encore lorsqu'un « produit » est dû à plusieurs services publics. Le calcul des contributions relatives s'avère alors très délicat. La question de la prise en compte de la qualité est également abordée. Différentes solutions sont envisageables sur la base d'une approche au cas par cas, mesurant la qualité par la différenciation du service, la réussite de l'activité ou la contribution apportée à un *outcome*. Les problèmes liés aux décalages dans le temps de mesure de la production sont aussi examinés.

Enfin, la mesure de la productivité est abordée. Compte tenu de ce qui a été dit auparavant, il paraît évident que la mesure de la productivité des administrations publiques doit remplir un certain nombre de préconditions. Tout d'abord, les indicateurs d'*inputs* doivent être complets et correctement déflatés, tenant compte des modifications de qualité. Ensuite, les indicateurs d'*outputs* doivent répondre aux règles du principe De plus, même si les indicateurs d'*output* et d'*input* sont individuellement satisfaisants et leur marge d'erreur acceptable, il peut ne pas en être de même pour leur ratio. Plus généralement, un ratio – fût-il correctement construit – ne peut rendre

entièrement compte de la performance de services publics complexes ayant des objectifs multiples. Comme signalé en introduction, il y a une différence de nature entre les estimations issues des comptes nationaux et les indicateurs de performance utilisables pour la gestion des services publics.

7. Le chapitre 7

Il s'intéresse au processus de mise en œuvre des principes décrits ci-dessus. Il part des pratiques existantes de l'ONS et examine comment elles peuvent être adaptées pour s'accorder aux principes définis. Des recommandations en sont tirées concernant : les accords à passer entre l'ONS et les statisticiens des ministères pour assurer la disponibilité de séries statistiques fiables ; le besoin d'une information publique diffusée par l'ONS sur les nouvelles mesures directes de production et leur intégration dans les comptes nationaux ; la nécessité d'un audit externe tous les trois ans pour chacun des secteurs de dépenses dont la production est mesurée de manière directe ; la poursuite des travaux de l'ONS sur ces sujets ; la collaboration de l'ONS avec d'autres pays ; la concentration des travaux de l'ONS sur l'amélioration des mesures dans les secteurs déjà concernés avant d'étendre l'approche à d'autres dépenses.

8. Les chapitres 8 à 11

Les implications des réponses à ces questions sont ensuite détaillées et discutées pour différents secteurs de dépenses dans les chapitres 8 à 11 (secteurs de la santé, de l'éducation, de la sécurité intérieure, de la protection sociale). À la fin de chacun des chapitres figure un résumé des travaux en cours ainsi que des pistes à explorer.

Références bibliographiques

- Afonso A., L. Schuknecht et V. Tanzi V. (2003) : « Public Sector Efficiency: An International Comparison », *European Central Bank Working Paper*, n° 242, juillet.
- Afonso A. et M. St. Aubyn (2004) : « Non-Parametric Approaches to Education and Health Expenditure Efficiency in OECD Countries », *Technical University of Lisbon, Economics Working Paper*, n° 1/2004, janvier.
- Centre d'étude des revenus et des coûts (CERC) (1994) : *La productivité globale dans les services publics*, La Documentation française.
- Trosa S. (2004) : « L'évaluation des politiques publiques », *Les Notes de Benchmarking International, Institut de l'entreprise*, janvier.

Complément D

Retour sur les expériences d'évaluation en France

Martine Perbet

Direction générale du trésor et de la politique économique

Les divers dispositifs d'évaluation des politiques publiques mis en place dans les décennies précédentes, que ce soit pour l'évaluation *ex ante* (rationalisation des choix budgétaires en 1970 et étude d'impact des projets de textes gouvernementaux en 1995) ou pour l'évaluation *ex post* (premier et second dispositifs d'évaluation placés auprès du Commissariat général du Plan en 1990 et 1998) n'ont pas permis d'influencer la décision publique, ce qui était pourtant leur objectif essentiel. La première partie de cette note exposera les principales raisons de cet échec. Les parties suivantes reviendront plus en détail sur chacune de ces tentatives afin d'en mettre en lumière la logique propre. Ces parties visent essentiellement à permettre de comprendre en quoi « le ver était déjà dans le fruit » lors de l'instauration de ces dispositifs. La dernière partie fera le point sur les chantiers en cours.

1. L'effort n'a pas porté sur l'ensemble des leviers nécessaires à l'instauration d'un véritable dispositif d'évaluation au service de la décision publique

Un regard porté sur l'élaboration des politiques publiques montre que « décision » et « évaluation » évoluent dans des mondes parallèles. Il n'y a pas systématisation de l'évaluation en vue de la décision publique. Il existe ponctuellement des évaluations de politiques publiques à la demande de l'exécutif (évaluations régulières de mesures réalisées pour le MINEFI, évaluation du programme d'infrastructures de transport réalisé en décembre 2002

par l'IGF et le Conseil général des Ponts et Chaussées, bilan des grandes lois après cinq ans de mise en œuvre...), mais ces évaluations ne s'intègrent quasiment jamais dans un processus dynamique d'évaluation dont l'objectif serait d'appréhender la réelle valeur d'une politique pour en informer le Premier ministre. Trois « erreurs d'aiguillage » expliquent cette absence d'ancrage des décisions publiques dans un *corpus* d'évaluation.

1.1. L'analyse économique ne s'est pas imposée comme méthodologie de l'évaluation

La culture administrative française se conjugue mal avec la tradition anglo-saxonne de l'évaluation économique. L'évaluation a donc souvent été entendue en France comme une obligation d'étude, de consultation et de concertation sur un thème donné, à charge pour l'évaluateur de définir lui-même le format de son évaluation. Ce flou dans les attentes du commanditaire s'est traduit par une grande hétérogénéité des produits réalisés dans ce cadre. Au pire, l'évaluation a été réduite à un exposé des objectifs à atteindre et des moyens retenus pour y parvenir (études d'impact des projets de textes gouvernementaux), au mieux elle s'est révélée comme une compilation des connaissances et des points de vue (évaluation *ex post* dans l'orbite du Commissariat général du Plan). Le questionnement d'une politique (i.e. d'un programme ou d'une mesure...) n'a pas impliqué en France que les solutions alternatives soient systématiquement identifiées et qu'elles soient caractérisées par une évaluation formalisée de leurs coûts et avantages respectifs. Or cela constitue le cœur de l'évaluation dans les standards internationaux, aussi bien en ce qui concerne l'évaluation *ex post* qu'en ce qui concerne l'évaluation *ex ante* (OCDE et, plus récemment, Union européenne...).

1.2. La gouvernance de l'évaluation s'est vue accordée peu d'attention

L'évaluation est au cœur de conflits d'intérêts qui s'expriment au sein même de la structure gouvernementale. Chaque ministre souhaite en effet mettre en avant ses propres propositions, ce qui ne va pas forcément de pair avec la volonté de voir celles-ci passées au crible d'une évaluation très rigoureuse. Voici pour l'évaluation *ex ante*. En ce qui concerne l'évaluation *ex post* le problème est sensiblement le même, peu de ministres envisageant sereinement de voir décortiquer les avantages et inconvénients de leurs politiques passées. Dans ce contexte il est primordial d'organiser les procédures d'évaluations afin d'imposer l'évaluation comme support de la décision publique et d'organiser la coopération de tous les services autour de l'évaluation. Or, alors qu'à l'étranger de nombreux pays s'engageaient dans cette voie, notamment pour l'évaluation *ex ante* des politiques publiques, peu de progrès ont été faits dans ce domaine. Malgré des constats récurrents de

dysfonctionnements de l'interministérialité (cf. rapport Tron), aucune procédure n'est venue fixer les règles du jeu interministériel et poser comme principe premier que la réalisation d'une évaluation constituait un préalable à la discussion. L'évaluation *ex ante* est donc quasiment inexistante, le processus interministériel se bornant à être le théâtre d'affrontements dont la conclusion est l'arbitrage du Premier ministre. Cette situation pourrait évoluer si les propositions du rapport Lasserre⁽¹⁾ étaient retenues. Celui-ci préconise que toute nouvelle mesure d'une certaine importance soit examinée en réunion interministérielle sur la base d'une étude d'impact (réalisée par le ministre responsable de la proposition), cette étude s'affinerait progressivement en tenant compte des remarques formulées lors des réunions.

1.3. Le modèle recherché pour l'évaluation *ex post* a souvent été celui d'une instance indépendante ayant le monopole de l'évaluation

Cette préférence suppose implicitement qu'il suffirait de réunir les « bons » acteurs au sein d'une même structure pour qu'émerge un produit de qualité incontestable. Les prévisions macroéconomiques avaient été ainsi tout d'abord confiées à un oligopole administratif (INSEE et DP), avant que les pouvoirs publics ne soient amenés à faciliter la création des instituts indépendants de prévisions (OFCE et Rexecode) afin de rendre possible une libre confrontation des expertises. Cette réforme avait fait suite aux vives critiques de la société civile sur l'opacité de ces prévisions, critiques reprises par le rapport Lenoir⁽²⁾. Un consensus existe aujourd'hui sur le bien-fondé de cette ouverture des prévisions à la concurrence et les critiques portent maintenant sur l'accès aux outils de l'administration (données et modèles). Cette émergence du pluralisme dans un domaine régalien mérite attention au moment où se pose la question de la configuration du dispositif d'évaluation français : l'évaluation s'appuie sur des instruments statistico-économiques, pour autant, elle ne peut être assimilée à une science exacte. Dans ces conditions, mieux vaut tableer sur la confrontation d'évaluations réalisées par différentes institutions (comme le montrent les exemples des États-Unis et du Royaume-Uni), que sur les évaluations d'une institution en monopole servant indifféremment tous les décideurs potentiels (Parlement, exécutif...).

Ce cloisonnement entre décision publique et évaluation se nourrit sans doute de la crainte du décideur politique de se voir imposer des solutions par la technocratie. Le souci de ce dernier de préserver son libre arbitre se fonde toutefois sur une appréciation erronée de l'articulation entre « déci-

(1) Le rapport Lasserre sur la réforme des études d'impact (La Documentation française, mars 2004) propose ainsi un processus dynamique de réalisation des études d'impact des projets de textes gouvernementaux en relation avec la procédure interministérielle.

(2) Rapport commandé en 1979 par Valéry Giscard d'Estaing à R. Lenoir et B. Prot.

sion » et « instruction de la décision ». L'évaluation participe à l'instruction de la décision – elle vise à identifier les solutions les plus pertinentes à partir des savoirs disponibles⁽³⁾ – mais elle laisse toute latitude au décideur pour trancher. Celui-ci conserve donc ses prérogatives même si l'évaluation peut exercer une pression indirecte sur le décideur pour qu'il justifie son choix, notamment dans la situation où il retiendrait une solution différente de la solution de premier rang.

Les pouvoirs publics ont ouvert à intervalles réguliers des chantiers d'évaluation. Différents dispositifs ont vu le jour depuis le début des années soixantedix, ils se distinguent les uns des autres par « l'objet » évalué (dépense publique/réglementation/politique publique dans son ensemble) et par le calendrier de l'évaluation (évaluation *ex ante* ou *ex post*). L'évaluation *ex ante*, c'est-à-dire en amont de la décision politique, concerne plutôt des mesures ponctuelles (dépense publique/réglementation) qui peuvent faire l'objet d'évaluations systématiques. L'évaluation *ex post* porte essentiellement sur des politiques d'ensemble choisies en fonction des enjeux et des problèmes révélés. Au stade actuel qui est celui d'une remise à plat de l'ensemble du système d'évaluation des politiques publiques (cf. partie 6)⁽⁴⁾, force est de constater que ces dispositifs n'ont, dans l'ensemble, pas répondu aux attentes des pouvoirs publics. Pourquoi ? C'est ce qu'un retour sur ces différentes tentatives permet de mieux comprendre.

2. La rationalisation des choix budgétaires, centrée sur les dépenses publiques, constitue une préfiguration non aboutie de la LOLF (1970-1984)

La rationalisation des choix budgétaires visait à réaliser un budget de programmes et non de moyens, sur le modèle des démarches Planning Programming Budgeting System (PPBS) engagées à l'étranger et notamment aux États-Unis. Cette méthode proposée par R. Mac Namara, ministre de la Défense, consiste à fixer des objectifs de moyen terme à l'action publique, puis à établir des programmes d'actions pluriannuels et à les traduire en tranches budgétaires annuelles avec un suivi de leur exécution. La pertinence des actions engagée est appréhendée à partir du corpus micro-économique. La mise en œuvre de celui-ci s'avère toutefois délicate car les politiques à évaluer sont complexes, les données manquent, les outils d'évaluation sont parfois insuffisants. Ces difficultés conduisent à réformer le PPBS aux États-Unis mais il a cependant ouvert la voie à un processus d'approfondissement de l'évaluation institutionnelle, en liaison avec le monde universitaire, qui continue encore aujourd'hui.

(3) L'évaluation a surtout le mérite de discréditer les fausses « bonnes solutions » qui ne résistent pas à une étude technique correctement conduite.

(4) Les caractéristiques des principaux dispositifs d'évaluation (sauf ceux dédiés aux études d'impacts des projets de textes) sont présentés en annexe 1.

En France, la réforme est pilotée par la Direction du budget du ministère des Finances qui la fonde sur le volontariat : des incitations financières (crédits du chapitre 37-93) visent à stimuler des travaux d'évaluation au sein de l'administration, ces travaux sont suivis par une commission interministérielle RCB. La réforme ne débouche cependant pas sur une véritable optimisation des dépenses publiques car le cadre budgétaire n'évolue pas : les budgets de programmes réalisés dans le cadre de la RCB ne constituent qu'un appendice à la procédure budgétaire classique qui continue à se dérouler dans le cadre défini par l'ordonnance de 1959. Le Parlement, qui n'a pas été impliqué en amont dans la réforme, ne se saisit pas des études réalisées. Faute de débouchés, la RCB périclite et se réduit à des crédits d'études gérés par les « cellules RCB » créées au sein de chaque ministère. La RCB n'en a pas moins suscité un développement substantiel des études et recherches à caractère évaluatif au sein des administrations (500 études mises en chantier, dont quarante études lourdes à caractère interministériel, Concorde, RER, sécurité routière, périnatalité, etc.).

La rationalisation des choix budgétaires, née de la convergence entre une volonté du monde politique de progresser vers une bonne gestion publique et une appétence des ingénieurs-économistes de la sphère administrative pour le calcul économique⁽⁵⁾, échoue donc. La prédominance d'une vision tutélaire de la dépense publique par le ministère des Finances a sans doute freiné l'instauration d'une démarche de partenariat avec les ministères dépeniers (objectifs des politiques, indicateurs...) ainsi qu'avec le Parlement. De cette période est resté un certain scepticisme sur la possibilité d'établir un lien direct et contraignant entre le résultat des évaluations et les décisions budgétaires.

3. Il faut attendre la fin des années quatre-vingt pour voir apparaître une relance de l'évaluation des politiques dans le cadre du « renouveau du service public » impulsé par M. Rocard

Après cet échec, la France tarde à s'intéresser à nouveau à l'évaluation des politiques publiques alors en plein essor dans les autres démocraties occidentales, notamment dans les pays anglo-saxons (cf. encadré 1). À partir des années quatre-vingt plusieurs initiatives se développent, mais sans véritable cohérence entre elles. Différents organismes sectoriels d'évaluation sont créés dans l'orbite de l'exécutif, tel le Comité national d'évaluation des établissements publics d'enseignement supérieur, une des premières structures nationales d'évaluation. De son côté le pouvoir législatif se dote, en 1983, d'un Office parlementaire des choix technologiques et scientifiques.

(5) Cf. note de méthode de M. Lévy-Lambert « Appréciation de la rentabilité économique des investissements » sous le double timbre de la Direction de la comptabilité publique et de la Direction de la prévision, janvier 1968.

1. Démarrage de l'évaluation à l'étranger

Aux États-Unis, la fonction de contrôle est très tôt associée à l'idée d'une évaluation de l'efficacité avec la création, en 1921, du *General Accounting Office* (GAO). Au lendemain de la Seconde Guerre mondiale, la multiplication des programmes de Welfare suscite une mutation des méthodes du contrôle qui se traduit par un appel plus fréquent aux sciences sociales. Les travaux des autres pays peuvent être classés en deux vagues : alors que les pays de la première vague d'évaluation étaient généralement dirigés par des réformistes, en période de croissance (sociaux-démocrates en République fédérale d'Allemagne et en Suède, libéraux au Canada), la deuxième vague de l'évaluation a plutôt été le fait de gouvernements conservateurs dans des périodes moins favorables (par exemple États-Unis entre 1968 et 1974, puis après 1979 ; Royaume-Uni entre 1970 et 1974, puis après 1979 ; Danemark après 1982 ; Norvège de 1981 à 1986 ; Pays-Bas après 1982) ; dans ce cas, l'évaluation a été plus orientée par la perspective d'économies budgétaires.

Source : Portail Internet de l'évaluation des politiques publiques : www.evaluation.gouv.fr.

L'idée de forger un instrument national d'évaluation prend corps avec la constitution de l'évaluation comme un axe fort de la politique de « Renouveau du service public » tels qu'exposés dans la circulaire de février 1989 signée par Michel Rocard. Cette démarche de renouveau n'est toutefois pas dépourvue d'ambiguïté car elle vise à la fois à rationaliser l'administration (optimisation des moyens, déconcentration, responsabilisation) et à mieux intégrer la demande sociale (moderniser les rapports entre l'État et les usagers du service public). Cette tension entre des objectifs de nature différente est perceptible à la lecture des deux rapports qui paraissent sur l'évaluation, le rapport Deleau en 1986 et le rapport Viveret en 1989⁽⁶⁾. Le premier s'inscrit dans une approche *économiste* des politiques alors que le second privilégie l'organisation d'un débat social sur les politiques publiques. Autrement dit la valeur collective d'une politique se détermine, pour l'un en mesurant les effets de cette politique, pour l'autre en amenant les acteurs sociaux à un consensus. Cette seconde conception va l'emporter.

(6) Rapport demandé en 1988 par Michel Rocard, Premier ministre, à P. Viveret : L'évaluation des politiques et des actions publiques, Paris, La documentation française, 1989. La commande portait aussi sur la procédure d'évaluation du revenu minimum d'insertion... » (cf. lettre de mission). De fait cette loi va innover : pour la première fois en France, une loi est votée pour une durée déterminée (trois ans) et son renouvellement est conditionné aux résultats d'une véritable évaluation qui sera conduite par la Commission nationale d'évaluation du RMI (1989-1992).

La semi-indépendance du pôle d'évaluation créé par le décret du 22 janvier 1990 est originale dans le paysage international de l'évaluation. Ce pôle placé auprès du Commissariat général du Plan repose sur trois piliers : un Comité interministériel de l'évaluation (CIME), présidé par le Premier ministre, a vocation à arrimer l'évaluation à l'exécutif ; des instances d'évaluation ouvertes sur la société civile jouent le rôle de maîtres d'ouvrage des évaluations (rédaction des cahiers des charges, sélection des maîtres d'œuvre, responsabilité du rapport final) ; un Conseil scientifique de l'évaluation (CSE), composé de personnalités qualifiées, est « chargé de favoriser le développement des méthodes d'évaluation et de définir une déontologie en la matière »⁽⁷⁾. De fait ce dispositif se situe aux antipodes de la RCB aussi bien en ce qui concerne son positionnement institutionnel (Commissariat général du Plan au lieu du ministère des Finances), que son champ d'évaluation (grandes politiques au lieu de programmes), et que la temporalité des évaluations (évaluation *ex post* au lieu d'évaluation *ex ante*).

Le bilan de ce dispositif est très mitigé. La volonté d'associer de nombreux acteurs au sein des instances s'est traduite par la constitution de structures lourdes et rigides. Celles-ci ont finalement produit un nombre réduit de rapports (*cf.* encadré 1), souvent avec retard. Mauvais cadrage avec le temps politique, faible qualité opérationnelle de nombreux rapports, vont conduire à un désintérêt progressif du gouvernement. Le comité interministériel de l'évaluation ne se réunira d'ailleurs qu'à trois reprises (juillet 1990, mars 1991, mars 1993) et l'une des dispositions essentielles du décret de 1990, la délibération du Comité interministériel sur les suites à donner aux évaluations, restera tout simplement lettre morte.

La réforme de 1998 cherche une voie de sortie dans un nouveau positionnement institutionnel, plus éloigné de l'exécutif, mais ne réactive pas la réflexion sur la méthodologie. Ceci aurait été bien nécessaire, le CSE s'est en effet centré sur la gestion managériale des évaluations ce qui n'a pas permis de dégager des références communes en matière d'évaluation⁽⁸⁾. Ce sujet resurgit néanmoins au milieu des années quatre-vingt-dix à propos de l'évaluation des réglementations. La prolifération des normes amène en effet les pouvoirs publics à instaurer un examen systématique des textes gouvernementaux.

(7) *Cf.* deux textes de Michel Rocard : circulaire du 23 février 1989 sur le « renouveau du service public » et décret du 22 janvier 1990 relatif à l'évaluation des politiques publiques.

(8) *Cf.* annexe 2, résumé du « Petit guide de l'évaluation des politiques publiques » publié par le CSE en 1996 à la Documentation française.

2. Évaluations lancées dans le cadre de la procédure de 1990

Petite phrase de présentation de la suite :

- l'impact du développement de l'informatique sur l'efficacité de l'administration (1992) ;
- la politique de réhabilitation du logement social (1993) ;
- l'accueil des populations défavorisées dans certains services publics (1993) ;
- les politiques d'insertion sociale des jeunes en difficulté (1993) ;
- la politique d'aménagement des rythmes de vie de l'enfant (1994) ;
- la protection des zones humides (1994) ;
- la politique sociale et culturelle en faveur des agents de l'État (1995) ;
- la loi quinquennale pour l'emploi (1996) ;
- la prévention des risques naturels majeurs (1997) ;
- la politique de maîtrise de l'énergie (1997) ;
- la politique de la montagne (1998) ;
- la loi du 10 janvier 1991 relative à la lutte contre le tabagisme et l'alcoolisme (dite « Loi Évin ») ;
- les aides au logement (1993) ;
- la politique d'aide à la localisation des activités dans les zones de conversion (1994) ;
- la lutte contre la grande pauvreté (1995) ;
- les mesures en faveur des salariés âgés de plus de 55 ans (1996).

4. Après les dépenses budgétaires et les politiques interministérielles ce sont les réglementations qui sont soumises à évaluation par le biais des études d'impacts

Cette démarche repose sur une volonté politique de stimuler l'évaluation au sein de l'administration. L'insécurité juridique va croissant, au fur et à mesure de l'empilement des normes, l'évaluation *ex ante* des textes constitue l'instrument incontournable pour freiner et rationaliser l'action normative du gouvernement et de l'administration. Une expérimentation des études d'impacts est décrétée en 1995, et pérennisée en 1998⁽⁹⁾. Les circulaires appellent à la coopération des services, demandent un « bilan coûts-

(9) Cf. circulaire d'Alain Juppé en date du 21 novembre 1995 relative à l'expérimentation d'une étude d'impact et circulaire de Lionel Jospin du 26 janvier 1998 relative à l'étude d'impact. Cette dernière élargit le champ des textes devant faire l'objet d'une étude d'impact (lois, PLF, décrets...).

avantages » des propositions de textes gouvernementaux, ébauchent un véritable système de gouvernance de l'évaluation (responsabilisation du ministre, envoi de l'étude d'impact aux ministres deux semaines avant la RI, enrichissement de l'étude d'impact au cours du processus interministériel, notification au dossier des divergences entre ministres, possibilité pour le Premier ministre de décider d'une contre-expertise, transmission de l'étude d'impact finale au Parlement)... Cependant, malgré ces points forts, les études d'impact vont dériver vers un simple exposé des motifs qui sera rédigé à la hâte par le service initiateur de la proposition. Pourquoi ?

Essentiellement parce qu'il ne suffit pas de poser des principes, il faut aussi indiquer la façon de les mettre en œuvre. Cette nécessaire articulation entre les concepts et la pratique apparaît dans des domaines connexes. Le premier président de la Cour de cassation, Guy Canivet, a ainsi récemment souligné la difficulté d'intégrer les enjeux économiques dans les décisions juridiques sur la base des seuls principes (par exemple « réparer l'intégralité du préjudice subi, sans qu'il en résulte ni perte, ni profit pour la victime » ; « octroi de réparation pour préjudice »...). Il a donc demandé que soient définies de véritables règles d'évaluation (comment estimer le préjudice ? par rapport à quelle situation de référence ?), qui seules peuvent assurer une véritable cohérence des jugements⁽¹⁰⁾.

Inscrire effectivement les principes dans la réalité suppose de préciser les méthodes et de les rendre accessibles aux évaluateurs potentiels. Ceci concerne aussi bien les techniques d'évaluation (manuel de calcul économique appliqué) que les procédures encadrant la mise en œuvre de l'évaluation (par exemple rejet par le Premier ministre des propositions ne s'accompagnant pas d'une étude d'impact en bonne et due forme). Il faut également savoir mobiliser les compétences afin de créer en amont le socle de l'évaluation (instruments, pôle de compétences) et de réaliser les évaluations en aval. Le modèle développé par différents pays, dont le Royaume-Uni et les États-Unis, correspond en général à un réseau d'évaluation constitué d'une « tête » (services du Premier ministre) et de ramifications dans les services⁽¹¹⁾.

Le dispositif d'études d'impact est actuellement en cours de révision. Les études se sont en effet progressivement vidées de leur sens. Ce phénomène s'est amplifié avec l'accroissement par la circulaire de 1998 du nombre de textes devant faire l'objet d'une étude d'impact (lois, PLF, décrets...). Les services ont été potentiellement submergés. Un nouveau dispositif devrait donc être défini prochainement (*cf.* partie 6).

(10) *Cf.* articles *Les Échos*, 1^{er} mars et 10 mai 2004 en annexe 4.

(11) *Cf.* notes D3-04-009, « Études d'impact : exemples étrangers » et D3-04-058 « La décision publique au Royaume-Uni ».

5. En 1998 le dispositif d'évaluation des politiques évolue mais aucun changement majeur n'apparaît

Face à l'échec du CIME comme relais entre évaluation et décision, les pouvoirs publics choisissent de confier l'évaluation à une structure indépendante (CNE), à charge pour les pouvoirs publics d'intervenir en amont et en aval de celle-ci. Le Conseil national de l'évaluation est composé de personnalités désignées par le Conseil économique et social, le Conseil d'État et la Cour des comptes, de personnalités qualifiées reconnues pour leurs compétences (en matière d'évaluation et de sciences économiques, sociales ou juridiques) et d'élus représentant les trois principales associations (Association des maires de France, Association des départements de France, Association des régions de France), innovation qui devait permettre de prendre en compte la territorialisation de plus en plus affirmée des politiques publiques. De nouvelles règles tendent à rendre la procédure plus compatible avec les processus de la décision publique (programmation, modes de choix, délais des travaux, etc.). Il est mis fin aux fonctions du Conseil scientifique de l'évaluation⁽¹²⁾.

Afin d'insérer les évaluations dans le processus décisionnel, le Premier ministre arrête le programme d'évaluation sur la base des thèmes proposés par les ministres et, en aval, le Commissariat général du Plan définit les suites à donner aux évaluations et les inscrit « dans les procédures de décision ministérielles et interministérielles » (circulaire du 28 décembre 1998). Le Commissariat général du Plan voit d'ailleurs son rôle affirmé dans la procédure : secrétaire du Conseil national de l'évaluation, il doit aider au montage de projets, il est aussi appelé à assurer le développement de l'évaluation dans l'administration française et, pour cela, pilote un réseau de l'évaluation constitué par l'ensemble des « hauts fonctionnaires de l'évaluation » nommés dans les ministères.

Le bilan de ce dispositif est toutefois peu différent du précédent. Le pilotage du dispositif par une structure indépendante (CNE) n'améliore pas fondamentalement l'évaluation des politiques publiques. Les rapports produits sont plus nombreux que dans le précédent dispositif, mais ils restent encore volumineux et de qualité variable. Leur publication permet une meilleure information des citoyens mais a, en pratique, peu d'impact sur la vie politique. L'ancrage des évaluations dans les procédures de décision, comme la diffusion de la culture d'évaluation dans la sphère administrative, ne progressent pas. Les nouvelles missions du Commissariat général du Plan ne prennent pas corps, car il y avait sans doute une illusion majeure à penser que le CGP puisse faire du pilotage interministériel de l'évaluation à cadre institutionnel inchangé.

(12) Cf. décret du 18 novembre 1998, précisé par la circulaire du Premier ministre du 28 décembre 1998.

L'échec de la tentative précédente pour fonder les études d'impacts sur l'analyse économique semble avoir eu pour effet un rejet des évaluations économiques. Si le décret du 18 novembre 1998 indique que *l'évaluation interministérielle a pour objet d'apprécier l'efficacité de cette politique en comparant ses résultats aux objectifs assignés et aux moyens mis en œuvre*, il ne précise pas les méthodes à utiliser pour porter cette appréciation. Il indique même plus loin que *les projets dont les objectifs relèvent davantage de la recherche en sciences économiques et sociales que de la gestion des affaires publiques seront orientés vers d'autres dispositifs*, ce qui suggère la crainte d'une dérive vers des évaluations trop techniques. Se confirme ainsi le clivage entre « l'évaluation à la française », utilisant des méthodes hétérogènes, et « l'évaluation anglo-saxonne », fondée sur l'analyse économique.

6. Les chantiers en cours

L'évaluation des politiques publiques est en cours de rénovation. Deux démarches se déroulent concomitamment.

L'une concerne l'évaluation *ex post* des politiques publiques. Le Commissariat général du Plan avait été chargé en 1998 d'assurer le secrétariat du Conseil national de l'évaluation (CNE) et de contribuer au développement de l'évaluation dans l'administration (décret du 18 novembre 1998 et circulaire du Premier ministre du 28 décembre de la même année). Le CNE a été installé le 16 février 1999. Entre 1999, date d'installation du CNE, et 2002, la présidence de celui-ci a été assurée par Yves Cousquer. Le renouvellement des membres du CNE n'est pas intervenu au terme du mandat de trois ans. Depuis cette date, le Commissariat général du Plan a mené à leur terme les évaluations décidées antérieurement. Des réunions interministérielles se sont déroulées fin 2003 et courant 2004 afin de mettre en place un dispositif prenant la suite du Conseil national de l'évaluation (CNE). Il ne serait sans doute pas placé sous l'égide du Commissariat général du Plan car, dans sa conférence de presse du 24 septembre 2003, « Transformation du Plan : mission, objectifs, organisation », le Commissaire au Plan a indiqué qu'il avait proposé au Premier ministre que le Commissariat général du Plan se recentre désormais sur la fonction prospective et renonce au rôle qui était jusque-là le sien en matière d'évaluation des politiques publiques.

L'autre concerne le *dispositif d'études d'impact* qui est actuellement en cours de révision⁽¹³⁾. La dérive des études réalisées dans le cadre de 1998 ayant conduit à suspendre l'application de la circulaire pour une durée de dix-huit mois à compter du 1^{er} septembre 2003, les études en cours sont

(13) Cf. les deux circulaires du Premier ministre signées en 2003 qui visent à mettre en ordre de marche l'administration au service d'une réglementation plus efficace (haut fonctionnaire responsable de la qualité réglementaire, charte, cadre transitoire pour les études d'impacts du 1^{er} septembre 2003 au 1^{er} mars 2005).

guidées par des prescriptions temporaires. Un nouveau texte devait être publié d'ici le 1^{er} mars 2005. Il devrait s'inspirer du rapport rendu en mars 2004 par le groupe de travail du Conseil d'État (présidé par Bruno Lasserre). Ce rapport préconise de concentrer les études sur les textes porteurs d'enjeux importants, il met l'accent sur la nécessité de réaliser des bilans coûts-avantages des réglementations en s'appuyant sur les outils économiques et de faire valider les études par un pôle de compétence.

Conclusion

Par rapport à d'autres pays, notamment anglo-saxons, l'organisation de l'évaluation a été tardive en France, et elle a aussi emprunté une voie originale à plusieurs égards : elle s'est focalisée sur l'évaluation *ex post* des politiques interministérielles alors qu'à l'étranger les efforts portaient principalement sur l'évaluation *ex ante* des projets gouvernementaux ; elle s'est voulue collégiale réunissant exécutif, parlement, société civile autour d'un même diagnostic ; elle a mis l'accent sur la pluridisciplinarité alors que l'économie émergeait comme standard international de l'évaluation.

De fait les options retenues jusqu'ici en matière d'évaluation des politiques publiques n'ont pas débouché sur des évaluations utiles à la décision publique : faible qualité opérationnelle de nombreux rapports d'évaluation, mauvais cadrage de ceux-ci avec le temps politique, inefficacité de la régulation scientifique : le constat est établi en 1995 par le rapport Tenzer⁽¹⁴⁾ à propos de l'évaluation *ex post* reste d'actualité. L'évaluation *ex ante* s'est, elle, réduite comme une peau de chagrin les études d'impact ayant dérivé vers un simple exposé des motifs des textes gouvernementaux. L'expérience RCB d'un pilotage de la dépense publique fondé sur l'évaluation de programmes, la première tentative d'implémenter l'évaluation en France, a échoué faute d'une modification de l'ordonnance de 1959. Cette modification est intervenue en 2001 avec la LOLF ouvrant la voie à cette réforme.

Trois « erreurs d'aiguillage » expliquent cette absence d'ancrage des décisions publiques dans un corpus d'évaluation : l'analyse économique ne s'est pas imposée comme méthodologie de l'évaluation, ce qui s'est traduit par une grande hétérogénéité des études produites ; l'absence de gouvernance de l'évaluation (incitation, contrôle, sanction) n'a pas permis de mobiliser au mieux les ressources de l'administration ; la préférence des pouvoirs publics pour un monopole d'évaluation a nuit au développement de celle-ci. La pause institutionnelle en matière d'évaluation peut toutefois représenter un avantage. Elle permet de tirer un bilan des expériences passées pour mieux définir une architecture rénovée de l'évaluation en France.

(14) Tenzer Nicolas (1995) : « *L'évaluation : une nécessité stratégique pour un État moderne* », *Rapport au Commissaire au Plan*, La Documentation française, janvier.

Complément E

La décision publique au Royaume-Uni

Martine Perbet

Direction générale du trésor et de la politique économique, MINEFI

Les réformes de l'État et des politiques publiques sont devenues des questions centrales, en raison de leur impact sur les finances publiques mais aussi en raison des répercussions sur la productivité globale des économies. Les réformes entreprises par le Royaume-Uni constituent un modèle de référence. Les aspects les plus connus concernent la réforme de l'État : redéfinition des frontières entre secteurs privé et public, transformation radicale des structures administratives avec la création d'agences chargées des prestations de services (antérieurement assumés par les ministères), avec aussi introduction dans la sphère administrative de modèles et de méthodes provenant du milieu des affaires. En revanche on connaît moins les mutations opérées au sein du gouvernement afin de moderniser la procédure d'élaboration des politiques publiques. C'est l'objet de ce complément.

La première partie met l'accent sur l'étendue des pouvoirs du Premier ministre, qui dispose de tous les leviers nécessaires pour réformer : il dispose d'une administration recentrée sur l'élaboration des politiques et il s'est doté d'une capacité d'expertise propre (*Cabinet Office*), très réactive, utilisant des experts d'origine variée. La deuxième partie s'attache au processus interministériel caractérisé par une circulation dynamique d'information entre services, sous la responsabilité d'un ministère chef de file. Une gouvernance pragmatique de la coordination interministérielle est mise en œuvre grâce à des règles procédurales et des référents méthodologiques disponibles sur le site Internet du Premier ministre. Sont ainsi bien mis en avant les principes d'évaluation que doivent appliquer les départements à

leurs propres propositions législatives... ainsi qu'à celles de l'Union européenne. La troisième partie s'intéresse aux mécanismes mis en œuvre par le gouvernement britannique pour influencer sur les décisions de l'Union européenne.

1. Le rôle du gouvernement britannique a été recentré sur l'élaboration des politiques publiques

1.1. Un exécutif resserré autour du Premier ministre

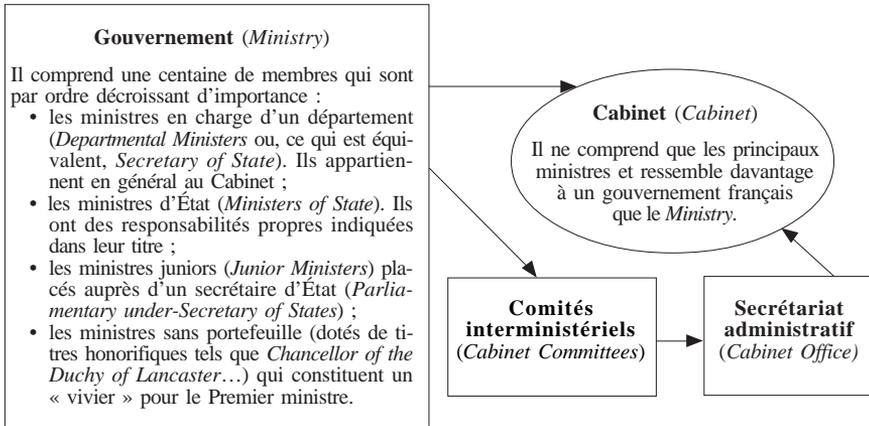
Le gouvernement (*Ministry*) conduit les politiques de la majorité parlementaire. Il joue en quelque sorte le rôle de « l'exécutif » du Parlement. Le Premier ministre est le chef du parti majoritaire à la Chambre des communes et les membres du gouvernement sont choisis au sein de ce parti. Le gouvernement agit à l'unanimité (responsabilité collective) mais chaque ministre est responsable de l'action de son ministère et doit en répondre devant le Parlement (responsabilité ministérielle).

Pour autant, ce *Ministry* composé d'une centaine de membres n'est pas vraiment l'instance gouvernementale. C'est le Cabinet (*Cabinet*), version réduite du *Ministry* puisqu'il ne comprend que les principaux ministres, qui ressemble davantage au Gouvernement français. Le Cabinet rend compte directement au Premier ministre.

Le Premier ministre exerce à la fois les fonctions d'un chef de gouvernement et celles d'un chef d'État. Il impulse les politiques, est responsable des engagements du gouvernement, et assure la communication du gouvernement auprès du Parlement et de l'opinion publique. Il nomme librement les membres du Cabinet et désigne les autres membres du gouvernement. Il nomme les directeurs et chefs de service des administrations. Il détient aussi le pouvoir de dissoudre la Chambre des Communes (même si ce pouvoir appartient formellement au souverain).

La figure suivante rappelle les principaux pôles du gouvernement : le *Ministry* dans lequel seuls les ministres en charge d'un département jouent un rôle important (*Departmental Ministers* ou, ce qui est équivalent, *Secretary of State*)... car ils appartiennent au Cabinet qui est la véritable instance gouvernementale. Le Cabinet s'appuie sur un Secrétariat administratif (*Cabinet Office*), véritable service, dont le rôle essentiel est de concevoir des politiques et de superviser leur mise en œuvre. Des structures transversales (*Cabinet Committees*), qui se réunissent au niveau des ministres, traitent de thèmes interministériels, parfois très larges comme l'environnement, l'emploi, ou au contraire très ciblés comme « Londres » ou « les importations illégales ». Ces comités sont rattachés au *Cabinet Office*.

Organisation du gouvernement



Source : Auteur.

1.2. Le Premier ministre dispose directement des principaux leviers de l'administration

L'organisation de l'administration a été profondément modifiée par le processus de réforme qui *s'est accéléré* à la suite du rapport « *Improving Management in Government: The Next Steps* » (Améliorer la gestion publique : prochaines étapes) publié en 1988 par une unité récemment créée du *Cabinet Office*. Ce rapport faisait essentiellement deux critiques au système administratif britannique : d'une part, la fonction publique était jugée trop nombreuse et diversifiée pour être administrée comme une seule entité et, d'autre part, les agents n'avaient que peu d'expérience en matière de gestion des services alors que c'était leur tâche principale. Ce constat a entraîné la mise en œuvre d'une approche nouvelle du fonctionnement du gouvernement, proche de celle qui prévaut dans tous les pays ayant opéré des réformes : le gouvernement demeure le maître d'œuvre de la détermination des politiques, de la réglementation et de l'allocation budgétaire, mais rien n'impose qu'il agisse directement en matière de gestion et d'administration de programmes et, *a fortiori*, qu'il demeure ou devienne un fournisseur de services à destination des usagers.

L'administration a donc été restructurée sur la base d'une spécialisation des missions. Elle comprend aujourd'hui :

- les départements ministériels qui déterminent les politiques, leurs domaines d'intervention sont à peu près semblables à ceux des ministères en France : éducation, transport, emploi, santé, commerce et industrie, défense, affaires étrangères, culture, environnement, agriculture (ces deux thèmes sont traités au sein du même département), justice (il n'y a pas à propre-

ment parler de ministère de la Justice : celui-ci est éclaté entre le *Department for Constitutional Affairs*, le *Crown Prosecution Service* et certaines parties du *Home Office*, une importante réforme en cours est justement la création d'un ministère regroupant ces différentes entités, sur le modèle français), et *HM Treasury* dont les compétences sont comparables à celles du MINEFI français⁽¹⁾ ;

- des départements non ministériels qui mettent en œuvre ces politiques (organismes de réglementation sectorielle par exemple) ;
- des agences d'exécution qui fournissent des services à la population ou des prestations pour les ministères, le gouvernement incitant dans ce cas le service acheteur à comparer les coûts du secteur privé à ceux du secteur public.

Au total l'élaboration des politiques est donc confiée à un « centre » qui comprend à la fois les départements ministériels et le *Cabinet office* sur lequel s'appuie le Premier ministre. La mise en œuvre des politiques, et surtout les prestations de services, sont situées à la périphérie. Cette distinction ne traduit pas un moindre intérêt pour ces tâches. Au contraire, un cadre général de régulation a été mis en place afin d'assurer l'efficacité des services produits (normes de qualité, engagement des agences, contrôles). Le « centre » chargé d'impulser les politiques s'est considérablement réduit puisqu'aujourd'hui les trois quarts des personnels de la fonction publique (déjà réduite d'un tiers depuis le milieu des années soixante-dix) travaillent dans les agences.

Le Premier ministre dispose donc d'une administration « noyau » pour impulser ses politiques. Il est également premier Lord du Trésor (*First Lord of the Treasury*) et ministre de la fonction publique (*Minister for the Civil Service*). Ces titres ont sans doute un caractère essentiellement honorifique car les fonctions correspondantes sont assurées respectivement par le *Chancellor of the Exchequer* et le Secrétaire du Cabinet mais cela vient néanmoins renforcer sa position prééminente vis-à-vis des autres ministres. Il a donc en main tous les leviers pour réformer l'État et les politiques publiques.

1.3. Le Premier ministre s'appuie sur une structure propre, le *Cabinet Office*, très réactive

Le Premier ministre travaille directement avec le *Cabinet Office*, structure polyvalente placée auprès de lui. Il se compose de trois services, l'un chargé des fonctions sensibles dévolues au Premier ministre (secrétariat

(1) Dirigé par le Chancellor of the Exchequer, second du gouvernement après le Premier ministre, ce département se compose de quatre services dotés de missions opérationnelles : budget et comptabilité publique (Chief secretary to the Treasury), fiscalité (Paymaster General), banque, services financiers et assurance (Financial Secretary), économie – notamment environnement, concurrence, R&D, exclusion – (Economic Secretary). Il dispose de différentes agences (gestion de la dette, épargne et investissement, statistiques nationales, commerce, monnaie...).

européen, secrétariat à la communication gouvernementale, secrétariat à la défense...), l'autre chargé de la gestion propre du *Cabinet Office*, et un troisième service, le plus important, dénommé *Building Capacity*. Ce service, particulièrement original au regard des structures françaises, est une structure de modernisation et de régulation publique, à l'interface des choix politiques et des moyens administratifs (cf. Commissariat général du Plan, 2003).

Dédié à la réforme de l'État et à la modernisation des politiques publiques, le *Building Capacity* joue le rôle de laboratoire d'idées au service du Premier ministre mais il exerce aussi des missions opérationnelles : il impulse des réformes à partir d'expérimentations, il suit des politiques, il produit des instruments (guides, procédures...), il assure la formation des cadres, il contrôle des analyses d'impacts de la réglementation... Les unités du *Building Capacity* (cf. encadré 1) sont idéalement placées dans le processus gouvernemental pour que leurs travaux puissent influencer sur les décisions.

1. Les cinq unités du *Building Capacity* du *Cabinet Office*

1. Deux unités concourent à la réforme de l'État

L'*Office of Public Services Reform* (OPSR – Agence de réforme des services publics) a été créé en 2001 pour aider le gouvernement dans ses projets de modernisation de l'administration (nationale et locale) et des services publics. Elle travaille sur les conditions de généralisation des expérimentations, met au point des « standards nationaux » et des « critères de qualité ». Elle regroupe trente personnes issues de l'administration mais aussi des secteurs public, privé et associatif.

Le *Corporate Development Group* (CDG – Groupe de gestion et de développement), unité de deux cents personnes, réfléchit aux questions de recrutement, formation, ressources humaines, gestion prévisionnelle des compétences, management, et plus généralement aux conditions et facteurs organisationnels de la réussite des projets de réforme. Il est responsable du *Centre for Management and Policy Studies* (CMPS), principal organisme de formation des cadres dirigeants des ministères.

2. Trois unités jouent un rôle majeur dans la rationalisation des politiques publiques

Le *Delivery Unit* (DU – unité d'aide à la décision) aide le gouvernement à définir ses priorités, en particulier en matière sociale, éducation, santé, délinquance, transport. Elle est chargée d'évaluer en temps réel le respect du plan de réforme des services publics, de conseiller et d'aider les ministères en cas de problèmes. La DU est constituée quarante personnes, de haut niveau, venant

du public et du privé. Elle s'appuie sur un réseau de correspondants connaissant bien le fonctionnement des secteurs public, privé et associatif.

La *Strategy Unit* (SU – unité stratégique) développe une réflexion globale sur les politiques publiques en s'attachant particulièrement aux problèmes transversaux et au long terme. Consacrée aux aspects mal pris en compte dans l'organisation traditionnelle des départements administratifs, cette structure d'une centaine de personnes travaille le plus souvent en collaboration avec les ministères sectoriels et développe des réflexions qui ont vocation à se traduire en actions publiques : création de nouvelles agences, refonte d'organisations, réaffectation de budgets, production de nouvelles normes ou législation, lancement d'expérimentations... Elle recherche des solutions pragmatiques et contrôlables (indicateurs d'évaluation, calendriers de mise en œuvre, choix de procédures pour observer les résultats).

La *Regulatory Impact Unit* (RIU – Unité d'évaluation de la réglementation) veille à la qualité de la réglementation. Elle édicte des principes de bonne réglementation et œuvre concrètement à l'émergence d'un système réglementaire qui soit compréhensible par les usagers. D'une part, elle a pour mission de toiletter le stock de réglementations (*Regulatory Reform Act* avril 2001) et, d'autre part, elle joue un rôle essentiel au niveau des propositions de réglementation susceptibles d'avoir des implications économiques et politiques importantes. Elle intervient, si besoin est, en appui lors de leur réalisation et examine les analyses d'impact de la réglementation réalisées par les services (législation nationale et législation de l'Union européenne). Autour de cette unité centrale gravitent des satellites, les Unités départementales chargées de l'impact de la réglementation (*Departmental Regulatory Impact Units* – DRIU) mises en place dans tous les départements de l'administration. Composées de une à quatre personnes, les DRIU assument les tâches courantes de coordination des activités de déréglementation et d'information des responsables de la réglementation. Une fonction de filtre «indépendant» leur est aussi attribuée pour qu'elles écartent les réglementations de médiocre qualité.

Une sixième entité, l'*Office of Government Commerce* (OGC – Agence d'achats publics) a été créée à l'origine pour améliorer les procédures d'achat de l'administration centrale. Elle a vu ses compétences récemment étendues : son responsable est à la tête du groupe pour la productivité du secteur public (*Efficiency Team*) et du programme d'évaluation correspondant (*Efficiency Review*). Elle travaille en étroite liaison avec le *Treasury*.

La volonté affichée de faire de ces unités des structures d'expertise de haut niveau, avec un recrutement diversifié (ministères, corps d'inspection, agences de développement, associations, « société civile », universités, et fondations, administrations étrangères, mais aussi entreprises privées), se heurte toutefois à quelques difficultés. Si le *Cabinet Office* et particulièrement le *Corporate Development Group* comportent une proportion de personnes très qualifiées supérieure à la normale, l'apport des personnes

recrutées du privé est en pratique assez limité, les différences culturelles constituant un véritable obstacle à l'intégration des nouveaux venus. Les fonctionnaires détachés des autres ministères sont le plus souvent des généralistes de bon niveau, mais plus rarement de véritables experts. Une preuve et une conséquence de cela sont le recours assez fréquent à des consultants externes, dont l'administration est devenue l'un des principaux marchés. De plus, la plupart de ces échanges sont de très courte durée (deux à trois ans) : les avantages sont un renouvellement constant des compétences, ainsi qu'une grande réactivité. L'inconvénient est la faible stabilité des équipes, avec les conséquences que cela implique en termes d'adaptation.

Enfin, au sein même des trois grandes fonctions du *Cabinet Office*, identifiées plus haut, chacune des unités jouit d'une autonomie assez grande, renforcée par l'organisation en projet. Le problème est que la communication interne n'est pas encore assez organisée pour éviter les doublons, les chevauchements voire les contradictions.

2. La procédure interministérielle, très encadrée, vise à assurer un fonctionnement optimal du dispositif d'élaboration des politiques

La spécialisation des départements ministériels rend indispensable une coordination préalable à la prise de décision, le point de vue du ministre initiateur de la politique devant être confronté à celui des autres ministres. Cette confrontation se situe à un double niveau, technique et politique. Technique dans le sens où l'analyse du dossier doit se nourrir des informations, études et suggestions recueillies dans l'ensemble de l'appareil administratif, politique dans le sens où le dossier arrivé à maturité peut réclamer un arbitrage.

2.1. Des instruments procéduraux et méthodologiques encadrent le processus de décision

La culture politique britannique se caractérise par une absence de formalisme et un esprit coopératif, ce qui se traduit par un processus interministériel « bien huilé ». Tous les projets de législation émanent du gouvernement car, bien qu'en théorie ce soit aux membres du Parlement de présenter les propositions de loi, ce sont en réalité les ministres du Cabinet qui en ont l'initiative. Les décisions sont prises lors de la réunion hebdomadaire du Cabinet. Elles sont préparées en amont par le *Cabinet Office* qui joue en cela un rôle comparable à celui du Secrétariat général du gouvernement en France.

En pratique la plupart des propositions de politiques se finalisent dans le cadre de comités interministériels (*Cabinet Committees*). Le Cabinet ne traite en effet que des principales politiques, c'est-à-dire celles d'import-

tance majeure ou celles ayant trait à des sujets particulièrement sensibles. Il peut lui arriver également de régler les conflits entre départements mais la tradition bien établie de dialogue entre services évite en général de tels arbitrages. Cette tradition est d'ailleurs relayée par des règles gouvernementales explicites qui encadrent le processus de décision.

Ces règles préconisent :

- un partage de l'information sous la responsabilité du ministère chef de file de la proposition. Celui-ci procède à de multiples consultations au sein du gouvernement et à l'extérieur de celui-ci avant de formaliser sa proposition. Cette concertation, menée à un stade précoce, permet de déboucher assez rapidement sur un projet consolidé qui se concrétise lors d'un comité interministériel (*Cabinet Committees*). La réunion de ce comité est préparée par le ministère chef de file auquel il revient d'obtenir l'accord formel des autres ministères. En cas de divergences persistantes le *Cabinet office* intervient pour faire un point « objectif » de l'état du dossier en direction du chef de file et assure alors le rôle de secrétaire du comité interministériel (rôle qui est en général délégué aux services du ministère chef de file). Cette intervention du *Cabinet office* se fonde sur le principe de subsidiarité ;

- une étude économique des différentes solutions. Les discussions au sein du gouvernement, ainsi que celles avec la société civile, s'effectuent sur la base de l'analyse d'impact (AIR ou RIA en anglais) préparée par le ministère concerné. Tous les départements de l'administration et toutes les agences proposant une réglementation importante doivent produire une analyse d'impact du type coûts-avantages pour la collectivité, en termes monétaires si possible. Cette analyse s'affine au fil de l'avancement du projet de texte (AIR préliminaire, partielle, complète, finale, cf. encadré 2 et le projet de loi n'est rédigé qu'une fois la proposition approuvée par le Cabinet.

- la référence à un cadre technique et procédural strict. L'action des services est guidée par une panoplie d'instruments disponibles sur le site Internet du *Cabinet office*. De nombreuses *checklists*, par exemple, rappellent les exigences gouvernementales en matière de procédures : une *checklist* rappelle au ministère initiateur de la proposition tous les éléments d'évaluation qui doivent être fournis à ses collègues avant la réunion du comité interministériel⁽²⁾. Une autre, la *Policy Makers Checklist*⁽³⁾, guide le service lors de la réalisation de l'étude d'impact de sa proposition : elle affiche une série de questions aidant le décideur à définir les domaines affectés par sa proposition puis le dirige automatiquement vers les manuels appropriés (bilan coûts-avantages, collecte des données, traitements statistiques, procédure de révision des évaluations, mais aussi par thèmes : développement durable, santé, droits de l'homme, etc.). *Checklists* et manuels sont rédigés de façon très pédagogique, afin d'être accessibles à l'ensemble des utilisateurs potentiels au sein des services (cf. encadré 3).

(2) <http://www.cabinetoffice.gov.uk>

(3) *Checklist* des décideurs politiques.

3. Policy Makers Checklist : quelques « trucs » pour guider vos évaluations

<http://www.cabinetoffice.gov.uk>

1. Prévoir les impacts de votre proposition

Voir *Livre vert du Trésor sur l'évaluation*.

2. Quelles sont les différentes options ?

Penser au scénario « Ne rien faire » et voir trois manuels :

- Un peu d'imagination pour une meilleure régulation ;
- Alternatives à la réglementation ;
- Livre vert du Trésor.

3. Quels sont les coûts et les bénéfices ?

Rappel des trois piliers du développement durable qui requièrent l'évaluation des impacts économiques, sociaux et environnementaux. Voir pour cela *Livre vert du Trésor* et d'autres références sur Internet en matière de santé.

L'évaluation doit prendre en compte le risque et pour cela il peut être nécessaire de disposer d'expertise scientifique. Sur ces deux points voir le *Livre orange du Trésor* (Gestion du risque. Une vue d'ensemble stratégique) et le document *L'expertise scientifique dans la décision publique*

4. Équité et justice

Évaluer l'impact sur des « populations sensibles » (femmes, jeunes, personnes âgées, minorités ethniques, zones rurales). Voir pour cela différentes références sur Internet.

Évaluer l'impact sur les petites entreprises. Voir le *Guide pour l'analyse d'impact*.

5. Avez vous consulté toutes les parties concernées ?

Voir le *Code sur les pratiques de consultation*

6. Avez vous mis en place des procédures pour évaluer votre politique ?

Voir le *Livre magenta : guide d'évaluation des politiques* (complément au *Livre vert* sur les questions de bilan *ex post* des politiques et de collecte de données en vue de ce bilan).

Parmi les manuels, deux constituent des références particulièrement importantes pour l'évaluation des politiques publiques : le manuel *Better Policy Making: A Guide to Regulatory Impact Assessment* (Améliorer le processus de décision publique : guide d'analyse de la réglementation d'impact) indique comment réaliser les analyses d'impact des propositions de réglementation, et le *Green Book: Appraisal and Evaluation in Central Government* (Livre vert de l'évaluation en administration centrale) constitue l'ouvrage de référence en matière d'évaluation économique.

Ce cadre n'est pas seulement théorique, les départements et agences réalisent ainsi près de 200 RIA chaque année. D'ampleur inégale, suivant en cela l'importance du texte, elles respectent globalement les prescriptions du manuel *Better Policy Making*. Elles sont souvent limitées par la disponibilité des données et parfois par le manque de compétences, mais ceci est inévitable dans la phase de démarrage d'un processus encore jeune. L'engagement du ministre responsable est matérialisé par sa signature précédée de la mention « J'ai lu l'analyse d'impact et je constate que les avantages sont supérieurs aux coûts ». La transparence est organisée par la publication de ces études sur le site Internet de la RIU. Une procédure d'expertise par le *National Audit Office* a été enclenchée pour la première fois en mars 2004. Son rapport fondé sur un échantillon de RIA réalisées au cours de l'année précédente, souligne la qualité des consultations engagées mais préconise diverses améliorations sur d'autres points, notamment par exemple de mieux rechercher des solutions alternatives à la réglementation.

Enfin diverses publications internationales mettent l'accent sur l'effort fourni par le Royaume-Uni pour alléger sa réglementation. Le rapport « *Doing Business in 2004: Understanding Regulation* » (La pratique des affaires en 2004 : comprendre la réglementation) de la Banque mondiale classe ainsi le Royaume-Uni parmi les dix pays comportant le moins de réglementations (avec l'Australie, le Danemark et les États-Unis). L'« Aperçu économique du Royaume-Uni » publié en janvier 2004, par l'OCDE indique de son côté que les régulations administratives et économiques du Royaume-Uni sont les plus faibles des pays membres en ce qui concernent les freins à la concurrence et les barrières au commerce.

2.2. La coordination interministérielle en France, sous l'égide du Secrétariat général du gouvernement, paraît beaucoup plus formelle

Il y a longtemps eu des réticences, en France, à l'existence d'une institution de coordination interministérielle placée auprès du Premier ministre, considérant que l'accord entre les ministres était un problème politique et non administratif (cf. Payre⁽⁴⁾, 2004). Le Secrétariat général du gouvernement (SGG) a néanmoins été créé à la Libération. Cependant, il ne constitue pas une institution de coordination interministérielle mais plutôt une structure administrative d'accompagnement du processus interministériel : son

(4) Maître de conférence de droit public.

travail consiste essentiellement à convoquer les réunions interministérielles (plus de mille réunions par an) et à officialiser les décisions du Premier ministre dans *les Bleus*. S'il veille au respect des règles du jeu constitutionnel, il ne joue aucun rôle dans le processus de choix des politiques. Il n'y a pas de fait de « règles du jeu interministériel » qui garantissent que les dossiers aient fait l'objet d'une étude technique robuste et que les informations aient été partagées avant d'être débattues en réunions interministérielles.

En pratique deux voies ont été explorées pour améliorer la coordination interministérielle :

- la nomination d'un ministre coordonnateur. Cette solution s'est avérée un échec car celui-ci n'a jamais disposé d'une autorité suffisante sur ses collègues qui jugeaient que sa fonction portait atteinte au principe d'égalité entre les membres du gouvernement ;
- la création de structures horizontales telles que la DATAR, la DGRST, la délégation à la ville. Ces structures de coordination ont fini par se « stratifier » et par s'intégrer à l'administration traditionnelle.

La multiplication des tentatives pour créer des instances transversales ou pour nommer des chefs de file, témoigne de l'inadaptation du dispositif des réunions interministérielles : organisées souvent dans l'urgence, l'information est mal partagée et manque en général pour ancrer la discussion interministérielle sur des éléments techniques robustes (*cf.* rapport Tron, Commission des finances de l'Assemblée nationale, 2003).

3. Une attention particulière est portée par le *Cabinet Office* aux politiques émanant de l'Union européenne

Comme dans les autres États membres, environ la moitié des nouvelles réglementations mises en application au Royaume-Uni découle d'initiatives de l'Union européenne. L'amélioration de la qualité de ces réglementations est considérée comme primordiale par le gouvernement britannique qui s'est attaché à optimiser son action au niveau européen. Le Royaume-Uni a ainsi mis en place un système administratif très efficace pour représenter et défendre ses intérêts à Bruxelles. À travers le double mécanisme du secrétariat européen du Cabinet Office – chargé de coordonner l'action gouvernementale – et de la représentation permanente à Bruxelles (*United Kingdom Representation* ou UKREP), le Royaume-Uni présente à Bruxelles des positions politiques soigneusement élaborées.

Geneviève Bertrand (2002) souligne à cet égard que le travail gouvernemental pour les affaires européennes est coordonné d'une manière très sensiblement différente de celle adoptée en France. Tout d'abord, l'*European Secretariat* se concentre sur l'examen des questions de fond du débat européen, en cours ou à venir, ce qui n'entre pas dans les missions du SGCI. Ensuite, autre point substantiel de différence entre le Royaume-Uni et la France, une part importante de l'étude des dossiers communautaires se fait

à Bruxelles, où la représentation britannique (*UKREP*) est composée de 63 cadres (chiffre similaire à la représentation française). Très mobiles, ses membres, tout acquis à la mentalité du démarchage commercial, d'enquête et d'information au plus proche de la conception des politiques et des textes, passent un temps non négligeable dans les bureaux de la Commission et effectuent d'incessants déplacements à Londres, évitant toute coupure avec les administrations centrales, et se nourrissant des échanges et des informations venant de ces dernières.

De plus, tandis que l'ambassadeur, représentant permanent du Royaume-Uni à Bruxelles, participe tous les vendredis matins à Londres à une réunion générale sur le programme de la semaine à venir, son collègue français n'effectue en général le déplacement à Paris qu'une fois par mois, pour participer à la préparation du Conseil des affaires générales et une fois par semestre pour la préparation rapprochée du Conseil européen semestriel. Enfin, à la différence des Français, les Britanniques ont la pratique d'une diffusion de l'information dont est responsable le ministère « chef de file » sur chaque dossier. Cette tradition de transparence conduit chaque fonctionnaire à communiquer systématiquement et spontanément toute information utile aux départements potentiellement intéressés. Qui plus est, les fonctionnaires britanniques des différents services de la Commission maintiennent le contact entre eux ainsi qu'avec leur représentation à Bruxelles.

Cette action concertée pour agir au mieux au niveau européen s'appuie, elle aussi, sur des instruments pédagogiques à destination des acteurs administratifs. Ainsi le manuel *Better Policy Making: A Guide to Regulatory Impact Assessment* cité précédemment, comporte un important volet consacré aux projets de textes européens : il formule des conseils à l'intention des responsables politiques et des responsables de l'administration et des conseillers juridiques concernés par l'élaboration de la législation de l'Union européenne et sa mise en application ; il indique également comment réaliser les AIR des projets de textes européens.

Conclusion

Au total la culture politique britannique se caractérise par une absence de formalisme et un esprit coopératif qui se traduit par une facilité « systémique » à rénover son processus de décision publique et à mettre en place de nouvelles institutions. À l'évidence le régime politique et la structure gouvernementale favorisent la convergence des volontés ministérielles autour d'un projet commun piloté par le Premier ministre. Cette cohérence est cependant renforcée par les réformes récentes qui visent toutes à s'assurer que les décisions (y compris celles émanant de l'Union européenne) sont bien prises dans l'intérêt de la collectivité.

Ces réformes engagées depuis une décennie, et pour certaines d'entre elles plus longtemps, ont amené à considérer de façon attentive tout projet

de politique publique et, pour cela, à organiser le processus interministériel de façon à responsabiliser le ministère initiateur. La discussion entre services, comme la discussion avec la société civile est ainsi engagée sur des bases techniques. Les mesures prises au Royaume-Uni s'inspirent pour l'essentiel des « bonnes pratiques » universellement reconnues (responsabilisation, *guide-lines*) par les institutions internationales (OCDE, Banque mondiale...). Des enseignements utiles peuvent en être tirés aussi bien pour nos propres processus de réforme que pour anticiper l'évolution des dossiers européens concernant l'évaluation des politiques publiques (travaux du groupe « Mieux légiférer », par exemple).

Références bibliographiques

- Bertrand G. (2002) : « La prise de décision dans l'Union européenne », La Documentation française, Coll. 'Réflexe Europe', 2^e éd., juillet.
Résumé disponible sur www.diploweb.com
- Commissariat général du Plan (2003) : « Les ressources prospectives de la décision publique au Royaume-Uni », *Les notes d'ALEPH*, n° 2, septembre.
Disponible sur <http://www.plan.gouv.fr/publications/fiche.php?id=444>
- Commission des finances de l'Assemblée nationale : (2003) : « Évaluation et prospective : quelle organisation ? », *Rapport de la mission d'évaluation et de contrôle*, Note n° D3-004-075, juin.
- National Audit Office (2004) : *Evaluation of Regulatory Impact Assessments Compendium Report 2003-04*, Press Notice.
Disponible sur <http://www.nao.org.uk/pn/03-04/0304358.htm>
- Payre J-P. (2004) : *Notion de science administrative*, Miméo, Université Pierre Mendès France, Grenoble 2.
- Banque mondiale (2004) : « *Doing Business in 2004: Understanding Regulation* ». Disponible sur rru.worldbank.org
- Cabinet Office (2003) : *Better Policy Making: A Guide to Regulatory Impact Assessment*. Disponible sur www.cabinetoffice.gov.uk
- HM Treasury (2002) : *Green Book: Appraisal and Evaluation in Central Government*. Disponible sur www.hm-treasury.gov.uk
- HMSO (Her Majesty Stationery Office Efficiency Unit) (1988) : *Improving Management in Government: The Next Steps*.
Disponible sur www.cabinetoffice.gov.uk
- OCDE (2002) : *Capacité du Gouvernement à produire des réglementations de grande qualité au Royaume-Uni*, OCDE mars.
- OCDE (2004) : *Étude économique du Royaume-Uni*, janvier.
Disponible sur www.oecd.org

Résumé

Ce rapport sur la réforme de l'État est axé sur l'analyse de la performance dans la gestion publique mais également sur celle des incitations à mettre en place pour orienter l'activité d'agents publics.

L'objectif poursuivi avec la nouvelle architecture budgétaire de la LOLF est de faire évoluer la gestion publique vers la performance, sur la base d'une stratégie, d'objectifs, et d'indicateurs précis pour guider et évaluer les politiques.

Le travail engagé avec la mise en place des indicateurs de la LOLF est donc essentiel. Il doit s'inscrire dans une perspective plus large, de construction d'un dispositif complet de « *reporting* » ou « *d'accountability* » des services et opérateurs, dont l'unité de base devrait d'abord être microéconomique. Ceci constitue, avec la nécessité de se défier des systèmes centralisés d'évaluation, aisément capturables par les groupes de pression et les *lobbys*, la proposition principale de cette partie du rapport.

1. Évaluation et efficacité

Les auteurs soulignent que l'évaluation des performances du secteur public est une question cruciale, car elle conditionne l'efficacité de l'action publique, les régulations publiques s'exerçant toujours en situation d'asymétrie d'information et le plus souvent hors de toute référence marchande (en particulier pour les fonctions régaliennes). Mais c'est aussi une question délicate. On sait en effet que, lorsqu'il s'agit de rémunérer des agents (ou un opérateur) dont les tâches sont multiples, des indicateurs partiels conduisent à une focalisation excessive, l'agent délaissant celles qui ne sont pas comptabilisées. Il faut donc résister à la tentation de trop utiliser des indicateurs qui vont générer ce type de biais, en distinguant bien, par exemple, l'évaluation des performances des services et l'évaluation individuelle, et en

s'attachant à prendre en compte tout ce qui contribue au potentiel de long terme des services ou de l'économie.

Certes, la multiplicité des tâches ou des mandants est souvent exagérée, ou mise en avant par ceux qui souhaitent échapper aux contraintes de performance. Mais c'est une réalité de la gestion publique, dont il faut donc bien connaître les conséquences : les objectifs de l'action publique sont complexes, multidimensionnels et souvent évolutifs (par exemple lors d'alternances politiques), notamment quant aux pondérations des différents critères ; les difficultés d'évaluation des performances dans ce contexte de tâches multiples sont renforcées, dès lors que l'on est privé de l'information transmise par les marchés ; l'activité du secteur public implique fondamentalement qu'un agent dépend de plusieurs mandants, dont les critères d'évaluation sont différents, si bien que tout mécanisme d'incitation qui ne serait pas conçu conjointement risque d'être d'une faible efficacité dans un contexte de tâches multiples.

2. Organiser l'évaluation

En effet, comme toute organisation, le secteur public est caractérisé par un ensemble de relations de type « principal-agent » dans lesquelles les asymétries d'information jouent un rôle essentiel pour concevoir les mécanismes d'incitations appropriés. La structure hiérarchique de l'administration crée, de plus, des possibilités de collusion entre les agents et leur superviseur. Ceci conduit, au-delà des recommandations habituelles, à progresser dans la mesure de la performance, notamment au niveau des services fournis (« *outcomes* »), à mettre l'accent sur les questions de méthodes et d'organisation de l'évaluation, en insistant sur la nécessité de s'assurer de la compétence des évaluateurs, de leur absence de conflits d'intérêts, et du besoin de « certification » du processus.

L'approche en termes d'évaluation par comparaison (« *benchmarking* ») est un moyen très utile dans cette perspective, qui doit être systématiquement développé. Lorsqu'elle est possible, la comparaison doit intégrer les modes de gestion privée. En effet, si les fondements de l'intervention publique demeurent inchangés par rapport à la vision qui s'était constituée après-guerre, leur mise en œuvre peut se concevoir dans différents types d'organisation, impliquant des frontières différentes entre public et privé, et ayant des implications variables en termes d'incitations.

Une attention particulière doit être portée enfin à l'évaluation de la qualité, à la fois parce que c'est une dimension essentielle des services publics, et parce que le débat public sur ce thème est un moyen puissant de stimuler la performance.

Il convient ainsi de compléter le travail engagé avec la LOLF sur les indicateurs au niveau de l'information du Parlement, en développant la me-

sure de la performance au niveau microéconomique des services, agences ou établissements publics. Celle-ci devrait impérativement renseigner, au-delà de l'information sur les moyens et l'activité, sur les résultats et services fournis au public, en permettant la comparaison avec des établissements opérant dans des conditions comparables (ou rendues telles). Ces résultats devraient être établis de manière indépendante, certifiés, et rendus publics. Sur la base de ceux-ci des propositions d'actions (« suites ») devraient systématiquement être présentées à leurs instances d'orientation, avec un dispositif de suivi.

Les auteurs insistent donc sur la nécessité de se prémunir contre les risques de capture de l'évaluation, ce qui conduit à écarter très nettement le modèle d'instance centralisée qui pourrait en avoir le monopole, et plus généralement sur les risques d'une vision trop technocratique de la mise en œuvre des démarches par objectifs dans le cadre de la LOLF, qui rencontrerait inévitablement les écueils des mécanismes de planification centralisée. En revanche, on souligne évidemment que le développement de la mesure de la performance conditionne la mise en ligne des objectifs des agents ou agences avec l'intérêt général.

Finalement, ce rapport met l'accent sur le fait que la mesure de la performance au niveau microéconomique ne peut porter ses fruits que si elle s'exerce dans un environnement général favorable. Ceci conduit à s'interroger sur les entités les plus pertinentes pour développer cette « *accountability* », sur la question des frontières des agences, du degré de discrétion laissé à celles-ci ou aux agents, de l'articulation des agendas du politique et de la gestion, des relations avec les responsables budgétaires (pluri-annualité) et sur le fonctionnement des marchés du travail internes à l'administration.

3. Commentaire

Philippe Mongin insiste sur la distinction à faire entre performance d'une part, objectifs et résultats d'autre part. Identifier la performance est difficile et il est tentant de confondre résultat et performance. Enfin, il est suggéré aux rapporteurs d'être plus précis s'agissant des recommandations finales du rapport, par ailleurs bien fondées, dans leur analyse.

Summary

Performance, Incentives and Public Sector Management

This report on state reform centres more on the analysis of public management as well as the incentives that need to be established in order to steer the activities of public sector agents.

The purpose of the LOLF's new budgetary architecture is to move public sector management towards performance, based on a precise strategy, objectives and indicators to guide and assess policies.

Efforts to implement LOLF indicators are therefore vitally important and must form part of a broader perspective – the construction of a complete accountability framework that covers services and operators, the base unit of which must be primarily microeconomic. This and the need to be wary of centralised evaluation systems, which are easily hijacked by pressure groups and lobbies, is the main proposal of this report.

1. Evaluation and effectiveness

The authors stress that performance evaluation in the public sector is a crucial issue because it determines the effectiveness of public action, bearing in mind that government regulations always operate in a state of asymmetric information and most often without any reference to the market (particularly for sovereign functions). However, public sector performance is also a sensitive issue. Indeed, there is growing awareness that when it comes to the remuneration of agents (or operators) who perform multiple tasks, partial indicators lead to an overly narrow focus on particular tasks, with the agent abandoning those that are not recorded. Therefore, it is important to resist the temptation to overuse indicators that tend to generate this type of bias, distinguishing, for example, between measures of service performance and individual appraisals, striving to incorporate everything that contributes to the long-term potential of services or the economy.

Certainly, the multiplicity of tasks and principals is often exaggerated or cited by those who wish to escape performance constraints. However, it is a reality of public governance whose impact must be clearly understood. Public policy objectives are complex, multidimensional and often evolving (e.g. during political shifts), particularly in terms of the relative weighting of various criteria. The difficulty in assessing performance in this context of multiple tasks lies in the absence of market information. Public sector activity fundamentally implies that an agent reports to several principals, who have different evaluation criteria, so that any incentive mechanism that is not developed cooperatively risks being ineffective in a multi-task context.

2. Organising evaluation

Like any organisation, the public sector is characterised by a set of «principal-agent» type relationships, in which information asymmetries play an essential part in designing appropriate incentive mechanisms. Furthermore, the hierarchical structure of administration creates opportunities for collusion between agents and their supervisor. Beyond the usual recommendations, this allows for progress in performance measurement, particularly at the level of outcomes, and an emphasis on issues surrounding the methods and organisation of evaluation, in particular on the need to ensure that the evaluators are competent and free from conflicts of interest and that the evaluation process meets quality standards.

From this perspective, benchmarking is a very useful method and needs to be systematically developed. Whenever possible, comparisons should incorporate methods of private management. While government intervention has unchanged foundations since the post-war vision, it can conceivably be implemented in different ways, which will alter the boundaries between public and private and have variable implications in terms of incentives.

Special attention must be paid to quality evaluation, both because it is an essential dimension of public services and because public debate on this issue is a powerful means of stimulating performance.

Therefore, it is important to complement the work undertaken on indicators at the level of parliamentary information in the context of the LOLF, by developing a performance measure of services, agencies or public institutions at the microeconomic level. It is imperative that this measure provides information not only about resources and activities but also results and public services, comparing them to institutions operating in comparable conditions (or conditions that have been made comparable). These results must be independently established, certified and published. On the basis of these results, follow-up proposals should be systematically presented to the steering bodies, using a monitoring system.

The authors stress the need to guard against the monopolisation of evaluation – resulting in a firm rejection of the centralised authority model – and, more generally, the risk of an overly technocratic method of implementing the objective-based approach within the framework of the LOLF, which would inevitably come up against the pitfalls of centralised planning mechanisms. However, performance measures must obviously be developed in order to align agents' or agencies' objectives to the general interest.

Finally, the report emphasises the fact that performance measurement at the microeconomic level will only bear fruit if it is conducted in a favourable overall environment. This raises the issue of most appropriate entities to develop this accountability, as well as the boundaries of agencies, the degree of discretion given to agencies and agents, the coordination of policy and management agendas, relationships with budget managers (multi-year budgeting) and the operation of labour markets within the administration.

3. Remarks

Philippe Mongin notes that the report is vague on the distinction that needs to be made between performance on one hand and outcomes on the other. Identifying performance is difficult, and it is tempting to confuse outcome and performance. Finally, he suggests that while the authors' final recommendations are soundly argued, they need to be more precise.

PREMIER MINISTRE

Conseil d'Analyse Économique

66 rue de Bellechasse 75007 PARIS

Téléphone : 01 42 75 53 00

Télécopie : 01 42 75 51 27

Site Internet : www.cae.gouv.fr

Cellule permanente

Christian de Boissieu

Président délégué du Conseil d'analyse économique

Hervé Bonnaz

Secrétaire général

Laurent Flochel

Conseiller scientifique

Microéconomie

Jérôme Glachant

Conseiller scientifique

Macroéconomie

Théorie de la croissance

Marie Salignon

Conseiller scientifique

Macroéconomie

Économie du travail

Christine Carl

Chargée des publications et de la communication

01 42 75 77 47

christine.carl@pm.gouv.fr

Agnès Mouze

Chargée d'études documentaires

01 42 75 77 40

agnes.mouze@pm.gouv.fr

