

Commissariat Général du Plan

Observatoire Economique

4145

après le Premier ministre, chargé de la Fonction publique et du Plan



COMMISSARIAT GENERAL DU PLAN

*Rapport du groupe de travail
« Méthodes d'évaluation des politiques publiques »*

EVALUER LES POLITIQUES PUBLIQUES
Méthodes, Déontologie, Organisation

Direction Générale
et Stratégie des Transports

DOCUMENTATION

Réf. n° 4145

Michel Deleau, président du groupe de travail
Jean-Pierre Nioche
Philippe Penz
Robert Poinsard

MAI 1986

© La Documentation Française — Paris, 1986
ISBN : 2-11-001702-3

*Ministre d'Etat
Chargé du Plan
et de l'Aménagement du Territoire*

PARIS, le 10 décembre 1984

*Commissariat Général
du Plan*

Monsieur Michel Deleau
Sous-directeur à la Direction
de la prévision
Ministère de l'Economie, des
Finances et du Budget
151, rue Saint Honoré
75001 - Paris

Monsieur,

La loi définissant les moyens d'exécution du 9^e plan (loi n° 83-1180 du 24 décembre 1983) souligne la nécessité de *faire avancer les méthodes de planification*, tout particulièrement dans le domaine de *l'évaluation des politiques publiques* : « L'analyse des résultats des politiques engagées pendant le 9^e plan, résultats toujours complexes et d'une interprétation délicate, fera appel au *développement des méthodes d'évaluation des politiques publiques*. Il importe que les différents départements ministériels prennent des initiatives en ce sens, de façon que l'introduction de ces méthodes s'étende progressivement à l'ensemble de l'administration. Ces études devraient servir à réorienter en tant que de besoin les ressources budgétaires dont l'allocation optimale est nécessaire dans le contexte des finances publiques du 9^e plan ».

Les résultats de l'opération de rationalisation des choix budgétaires au cours des quinze dernières années ont contribué à rendre l'action publique plus transparente et efficace. Ils ont toutefois davantage concerné les phases amont, d'étude et de prospective, au détriment des phases aval, d'exécution et de mesure. La situation de la France en la matière apparaît en deçà de celle d'autres pays de niveau de développement comparable, où l'évaluation *ex post* des politiques publiques est pratiquée de manière plus systématique et approfondie, et sert à la fois d'instrument pour l'action des administrations et de référence pour le débat social.

Il nous est donc apparu nécessaire que soit engagée pour le cas français, une réflexion sur les instruments disponibles, ou à créer, en matière d'évaluation des politiques publiques et sur les problèmes pratiques de leur mise en œuvre. Cette réflexion sera conduite au sein d'un groupe de travail réuni sous l'égide du Commissariat général du plan, dont je vous remercie d'avoir bien voulu accepter la présidence.

Les travaux du groupe porteront donc essentiellement sur les méthodes d'évaluation *ex post*. Après avoir précisé le champ des politiques publiques envisagées et les types d'effets à évaluer, le groupe abordera trois thèmes principaux :

- un examen des expériences françaises et étrangères les plus significatives ;
- un recensement des instruments disponibles pour l'évaluation des politiques publiques et une analyse de leur champ d'application et des conditions techniques de leur mise en œuvre ;
- une discussion des principales questions d'organisation administrative ou politique qui paraissent déterminantes pour le développement de l'évaluation en France.

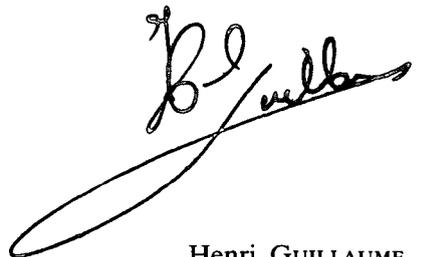
De manière plus précise, l'examen des expériences étrangères et françaises les plus significatives doit avoir pour fonction de recenser les causes de succès et d'échec des évaluations menées et d'apprécier les résultats atteints. Il visera à retirer des enseignements généraux sur les conditions de faisabilité technique de l'évaluation.

Compte tenu de cet examen, le groupe opérera un recensement des instruments d'analyse disponibles. Il devra s'attacher aux conditions techniques de leur mise en œuvre (par exemple le problème de l'information disponible, la référence ou non à une représentation explicite des mécanismes à l'œuvre, l'utilisation croisée d'instruments, la nécessité d'intégrer le processus d'évaluation dans la phase de définition même des politiques, etc.) et cherchera à identifier les situations types pour lesquelles une approche particulière (ou une combinaison d'approches) apparaît spécialement adaptée. Dans cette phase, le groupe devra établir des recommandations opératoires, utilisables dans les cas concrets.

Si le mandat du groupe est donc, tout d'abord, de fournir une réponse à la question : « comment faire une "bonne" évaluation des décisions publiques ? » il devra enfin identifier les points fondamentaux touchant à l'organisation économique et politique qui lui paraissent conditionner le développement de l'évaluation dans notre pays. Parmi ceux-ci figurent notamment l'alternative évaluation externe ou interne, le rapport entre contrôle et évaluation, l'unicité ou non des évaluations, l'accessibilité de leurs résultats.

Le groupe devra avoir terminé ses travaux à la fin de l'année 1985. Un rapport d'étape me sera remis en juin 1985.

Avec mes remerciements renouvelés, je vous prie d'agréer, Monsieur, l'expression de ma considération distinguée.



Henri GUILLAUME.

MINISTERE DE L'ECONOMIE,
DES FINANCES ET DE LA PRIVATISATION

DIRECTION DE LA PREVISION

Le Sous-Directeur

Paris, le 7 mai 1986

Monsieur le Commissaire,

Par lettre du 10 décembre 1984, vous avez bien voulu me confier la présidence d'un groupe de travail du Commissariat général du plan, chargé de préciser les conditions et les modalités d'un développement de l'évaluation des politiques publiques en France.

Je vous prie de bien vouloir trouver ci-joint le rapport qui présente l'analyse faite par le groupe et ses conclusions. Ce rapport traite successivement des divers points inscrits au mandat que vous nous aviez fixé. J'en résume, ci-après, les propositions essentielles.

*
**

Compte tenu de l'attention croissante portée, dans notre pays, à l'évaluation des politiques publiques, il a paru nécessaire, afin d'éviter de possibles ambiguïtés, de retenir une définition relativement stricte de celle-ci, qui permette d'en organiser la démarche, notamment au plan méthodologique, et de mieux marquer ses rapports avec divers genres voisins, tels que les contrôles administratifs. De la définition proposée et discutée dans la première partie du rapport (évaluer une politique publique, c'est reconnaître et mesurer ses effets propres), je souhaiterais souligner deux caractéristiques importantes.

La première a trait à la nature fondamentalement « extravertie » de la démarche évaluative, tournée vers l'extérieur de l'administration, puisqu'elle s'intéresse aux effets de l'action publique. Elle diffère, par là, des procédures traditionnelles, notamment de contrôle, largement internes et qui concernent la régularité de mise en œuvre des moyens plus que l'efficacité des résultats. On conçoit que l'évaluation puisse, ainsi, constituer un levier important pour l'adaptation des interventions publiques, de leurs domaines ou de leurs modalités.

La deuxième caractéristique importante réside dans la volonté de mesure quantitative. La reconduction sur longue période, sans justification réelle, de certaines politiques, par exemple en matière d'aides aux entreprises ou de protections réglementaires, a, sans nul doute, été facilitée par l'absence de toute évaluation quantitative, même sommaire, de leurs résultats effectifs.

Le mandat mettait l'accent sur les aspects méthodologiques. Le groupe s'est livré à un examen des méthodes existantes et des conditions de leur mise en œuvre (parties II et III).

A cet égard, la lenteur du développement de l'évaluation des politiques publiques dans notre pays ne peut être fondamentalement expliquée par la présence de

goulots d'étranglement techniques. Une panoplie d'instruments existe, leurs portées et limites ont fait l'objet de nombreuses analyses, même s'il est vrai que celles-ci n'ont longtemps attiré l'attention, en France, que d'un cercle restreint de spécialistes. Le rapport entend contribuer sur ce point à une plus large information. Cependant, par comparaison à la situation d'autres pays, les lacunes les plus flagrantes s'observent dans deux domaines.

D'abord, celui du recours à l'expérimentation, très peu développé en France. L'expérimentation, c'est-à-dire l'application d'une politique à un groupe, de taille restreinte mais représentatif, avec mesure des résultats avant généralisation éventuelle, ne saurait constituer une panacée. Mais, sans nul doute, un usage raisonné de cette démarche pragmatique permettrait d'éviter des déboires coûteux.

Second domaine où l'acquis est insuffisant : celui des outils aptes à fournir une représentation quantitative des comportements individuels et de la manière dont ils peuvent être influencés par les interventions publiques. L'absence de modèles microéconométriques, très développés dans certains pays, et corrélativement, l'usage parfois abusif des modèles macroéconométriques, conduisent, par exemple, à une appréciation exagérément comptable, ou biaisée, des interventions financières publiques en matière de fiscalité, d'aides ou de redistribution. Cette appréciation ne permet, en aucune manière, de conclure au bien fondé des politiques afférentes, notamment en ce qui concerne leurs effets sur moyenne période, ou d'indiquer les voies de réforme souhaitables.

Au-delà de ces aspects proprement techniques, le groupe a estimé nécessaire de proposer quelques règles déontologiques générales pour la conduite de l'évaluation. On peut les ramener à deux préceptes essentiels :

— la règle dite de **reproductibilité** qui a pour objet de garantir une transparence méthodologique des évaluations : l'affichage du protocole suivi par l'évaluateur doit permettre une reconstitution sans ambiguïté des conclusions obtenues. Cette règle est conforme à la volonté d'éclairage propre à la démarche évaluative.

— la règle dite de **séparation** qui institue une distinction entre, d'une part, les rôles de conception et d'exécution d'une politique et, d'autre part, le rôle d'évaluation. Il s'agit d'éviter les risques de biais ou d'auto-censure, inhérents à des pratiques trop centrées sur l'auto-évaluation. Cette règle, parfois contestée, paraît en définitive s'imposer, compte tenu des cloisonnements existant au sein de l'administration française, propices à la réalisation de ces risques.

De nombreux cas particuliers d'évaluation, français et étrangers, ont été étudiés par le groupe. Au-delà des spécificités apparentes, il importait de dégager quelques conclusions générales, par type de politiques (partie IV).

A cet égard, il ne ressort pas de spécificité sectorielle très marquée en matière d'évaluation. L'évaluation de l'efficacité, ou de l'inefficacité, d'aides aux entreprises soulève des problèmes analogues, qu'elles concernent les productions manufacturières ou agricoles. Une distinction peut par contre être introduite entre les politiques publiques assises sur la production de biens et de services, marchands ou non marchands, et les politiques publiques reposant sur des interventions financières ou réglementaires censées affecter le comportement des agents économiques.

Dans le premier cas, l'acquis des contrôles administratifs qui portent sur les

organisations responsables des productions concernées fournit souvent une base utile pour l'évaluation proprement dite. Les progrès passent par une amélioration des appareils de suivi statistique, un meilleur accès à l'information fine en matière d'entreprises publiques, la référence systématique à des groupes témoins permettant d'opérer des comparaisons continues et cohérentes. Il paraît par contre plus difficile, sauf pour les politiques très localisées, de se livrer à de véritables expérimentations.

L'acquis des contrôles administratifs traditionnels en matière d'interventions financières, et *a fortiori* réglementaires, est beaucoup moins mobilisable à des fins d'évaluation. En effet, la vérification de régularité de l'action publique est alors peu conclusive quant à son efficacité. Il importe donc de mettre un accent particulier sur l'évaluation de ces politiques qui ont conduit à des empilements largement opaques et d'une stabilité remarquable. Il faut, de ce point de vue, souligner les effets pervers d'une certaine organisation ministérielle de l'action publique qui, en privilégiant, assez naturellement, les relations verticales entre services administratifs et agents économiques, conduit à un fractionnement du champ d'intervention de l'Etat en une série de « monopoles bilatéraux », peu propices aux remises en cause. Ces effets pervers sont particulièrement forts dans le domaine réglementaire : les politiques afférentes, de coût financier direct nul (en apparence) pour l'Etat, ont peu de chances d'être réexaminées régulièrement lors de procédures d'arbitrage interministérielles de nature largement budgétaire. D'où la nécessité de mettre un accent particulier sur l'évaluation de ces politiques, ce qui appelle fréquemment que l'on dépasse les découpages ministériels traditionnels.

Enfin, la considération de diverses expériences étrangères (partie V) est d'un enseignement utile pour notre pays. Elle montre, d'une part, le retard de la France dans certains domaines, notamment en matière de diffusion de l'information sur les résultats des politiques et d'association des divers acteurs. Elle illustre, d'autre part, la nécessité d'un développement pluraliste de l'évaluation prenant en compte la spécificité des conditions nationales.

**

Au total, le développement de l'évaluation des politiques publiques dans notre pays apparaît comme une priorité impérieuse, notamment dans une perspective d'adaptation continue des interventions publiques et d'éclairage du dialogue social. Sa réalisation est à notre portée. Il serait toutefois illusoire de penser que ce développement puisse s'effectuer, dès à présent, de manière totalement spontanée sans un engagement soutenu de la puissance publique elle-même. Cette proposition ne signifie pas que l'évaluation des politiques publiques doive rester la chose de l'Etat. Il lui revient, par contre, d'assurer l'amarce d'un développement de l'évaluation où, à terme, il serait un acteur parmi d'autres.

Le groupe a rassemblé ses recommandations (partie VI du rapport) autour de cinq propositions principales : rendre les politiques publiques évaluables ; organiser les compétences ; renforcer l'évaluation des politiques publiques au sein de l'administration ; accroître la capacité d'évaluation externe à l'administration et favoriser sa diffusion dans la société ; identifier les thèmes prioritaires. Je ne reviendrai pas ici sur le détail de ces diverses propositions, mais souhaiterais en marquer les aspects qui m'apparaissent les plus significatifs :

- l'évaluation d'une politique est d'autant plus efficace qu'elle a été prévue et organisée dès la phase de formulation. Il y a donc un « réflexe évaluation » à développer très à l'amont, qui renvoie aussi bien à la mise en place de systèmes d'information *ad hoc* qu'à la définition d'échéances d'évaluation, avec possibilité d'infléchir ou de supprimer la politique si les résultats en sont négatifs. Le rapport fait également mention du fait nouveau, pour l'évaluation, que constitue la décentralisation. Le développement des compétences locales peut favoriser une pratique différenciée des interventions publiques, propice à l'éclairage des résultats. Une harmonisation des méthodes d'évaluation elles-mêmes paraît par contre souhaitable :
- le développement de l'évaluation ne signifie pas que la puissance publique doive mettre en place des structures spécifiques ou dégager des moyens nouveaux. La mobilisation soutenue de l'existant, en termes d'établissements de formation, de centres d'études, de moyens financiers, semble une voie efficace et adaptée. Il convient, en la matière, d'éviter toute dispersion, de manière à aboutir à la constitution de centres de compétence, notamment externes à l'administration, de taille suffisante ;
- l'évaluation ne progressera dans notre pays que si l'administration elle-même la développe fortement en son sein. Il faut toutefois rejeter l'idée d'une auto-évaluation généralisée des services par eux-mêmes : l'initiative de l'évaluation semble devoir relever de l'autorité ministérielle au même titre que l'initiative de la décision. En matière de publication des évaluations, il serait illusoire de recommander la diffusion absolument systématique des résultats et conclusions. En revanche, chaque opération d'évaluation peut être l'occasion d'une information publique factuelle, permettant aux divers partenaires sociaux de mener à bien leur propre évaluation s'ils le jugent utile. Enfin, le développement de l'évaluation au sein de l'administration semble devoir se conduire, de manière préférentielle, à un niveau interministériel à la fois pour améliorer la qualité des travaux en les soumettant à une plus large confrontation, favoriser la diffusion de leurs résultats au sein de l'administration et éviter la rétentation d'informations, permettre l'intégration de ces résultats dans les procédures d'arbitrage interministérielles, notamment de caractère budgétaire. Il va de soi que les différentes instances chargées de missions de contrôle et dont, d'ailleurs, l'activité en matière d'évaluation a sensiblement crû (Cour des comptes, Inspection générale des finances, Inspections ministérielles), devraient jouer un rôle particulier dans ce développement de l'évaluation des politiques publiques au sein de l'administration, étant entendu que la mobilisation de compétences techniques spécifiques, externes à l'administration, sera également nécessaire ;
- l'accroissement de la capacité d'évaluation externe à l'administration doit reposer tout d'abord sur le renforcement, évoqué plus haut, de pôles de compétence extérieurs, en suivant l'exemple réussi en matière d'expertise économique. Il doit conduire également à un rééquilibrage en faveur du législatif, beaucoup plus dépourvu de moyens d'évaluation que dans d'autres pays. Enfin, le groupe a estimé qu'une instance telle que le Commissariat général du plan pourrait jouer un rôle important, à la fois par sa vocation interministérielle, sa position charnière vis-à-vis des acteurs économiques et

sociaux, sa légitimité pour connaître des politiques « de moyen terme », appliquées sur longue période (par exemple sécurité routière, restructurations industrielles, environnement et aménagement du territoire, concurrence...), déterminantes pour le développement économique et social et dont l'évaluation ne relève pas naturellement ni par l'ampleur du champ concerné ni par l'horizon, des procédures traditionnelles ;

- l'évaluation des politiques publiques, enfin, ne saurait être conduite tous azimuts. Il paraîtrait efficace, dans un premier temps, de s'intéresser, pour reprendre les termes du rapport, aux politiques « silencieuses » et « enracinées » vis-à-vis desquelles la fonction de remise en cause de l'évaluation pourrait s'exercer efficacement. Cette observation concerne les politiques agissant par incitations réglementaires ou financières. J'ajoute également que les fonctions régaliennes traditionnelles (administration générale, défense, justice, sécurité) sont, bien plus en France que dans d'autres pays, préservées d'éclairages évaluatifs.

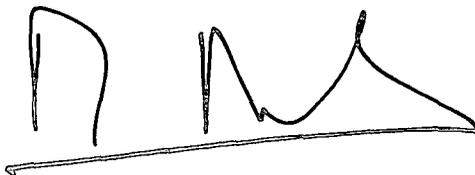
Le rapport n'avait pas comme objectif de fixer dans le détail, notamment institutionnel, la mise en œuvre de ces diverses recommandations. Elles ont paru, aux membres du groupe, susceptibles de garantir un développement solide et pluraliste de l'évaluation des politiques publiques, adapté à l'environnement national.

*
**

Je soulignerai, pour conclure, que le rapport qui vous est remis est le fruit d'un travail commun associant tous les membres du groupe, et qui a bénéficié des interventions de nombreux experts, dont la liste figure en annexe. La rédaction du rapport a été partagée avec Jean-Pierre Nioche, Philippe Penz, secrétaire du groupe, et Robert Poinsard.

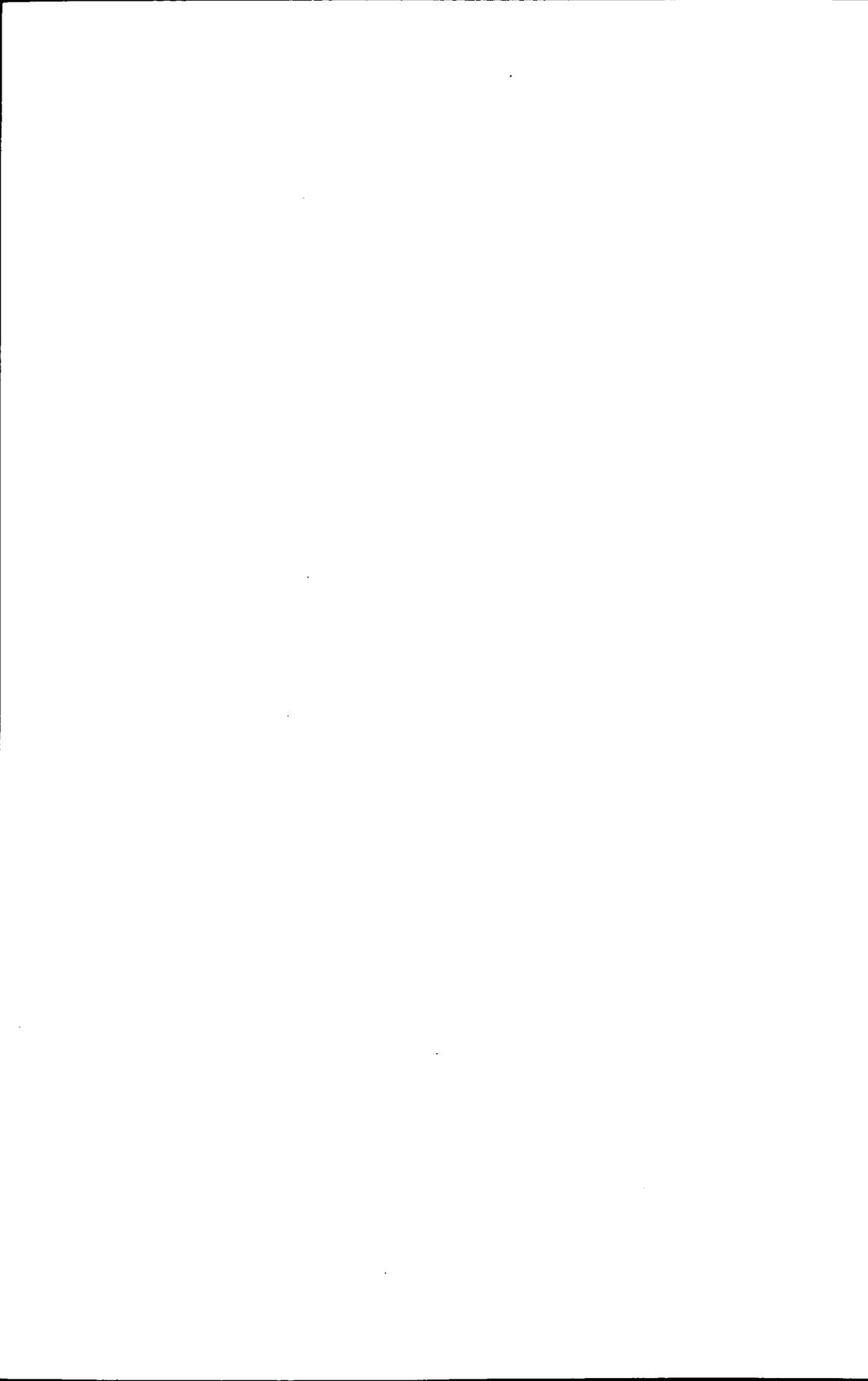
La diversité et la qualité de ces diverses contributions, dont je remercie ici les auteurs, me paraît également porter témoignage des chances existant actuellement pour un développement de l'évaluation des politiques publiques en France. Je souhaite que ce rapport, réalisé sous l'égide du Commissariat général du plan, y apporte une contribution efficace.

Je vous prie d'agréer, Monsieur le Commissaire, l'expression de mes sentiments distingués.



Monsieur Henri Guillaume
Commissaire au Plan
18, rue de Martignac
75007 Paris

M. DELEAU
Président du groupe de travail



Sommaire

Introduction

I — L'évaluation des politiques publiques : définition et frontières

I — 1 Les risques de l'imprécision	25
I — 2 Définition et attributs de l'évaluation	28
I — 3 Les genres voisins de l'évaluation : aide à la décision, contrôle de gestion, contrôles administratifs	35
I — 4 Évaluation des politiques et analyses de l'activité publique	43
I — 5 L'évaluation des politiques publiques : une approche spécifique	47

II — Les méthodes de l'évaluation

II — 1 Quelques éléments de terminologie	51
II — 2 L'évaluation sans groupe témoin	53
II — 3 L'évaluation par comparaison avec un groupe témoin non équivalent	63
II — 4 L'évaluation par comparaison avec un groupe témoin équivalent ou expérimentation vraie	71
II — 5 Évaluation et théorie : de l'étude de cas à la modélisation explicative	74
II — 6 Conclusion : la boîte à outils de l'évaluation	82

III — La mise en œuvre de l'évaluation

III — 1 Introduction	91
III — 2 La stratégie de l'évaluation	93
III — 3 Le choix d'un plan de recherche	101
III — 4 L'après-évaluation	105
III — 5 Pour une déontologie de l'évaluation des politiques publiques	106

IV — Les domaines de l'évaluation

IV — 1	Un découpage des politiques publiques	113
IV — 2	La production de biens et de services	118
IV — 3	Les interventions économiques financières : incitation, redistribution, taxation	123
IV — 4	La réglementation	125
IV — 5	La communication sociale	127
IV — 6	La réglementation macroéconomique	128

V — Les enseignements de quelques expériences nationales

V — 1	Évaluation et esprit des peuples	135
V — 2	L'exemple des États-Unis	136
V — 3	Autres expériences étrangères : R.F.A., Pays-Bas, Royaume-Uni, Suède, Suisse	140
V — 4	Le cas de la France	149
V — 5	Les limites d'un développement spontané de l'évaluation dans notre pays	153

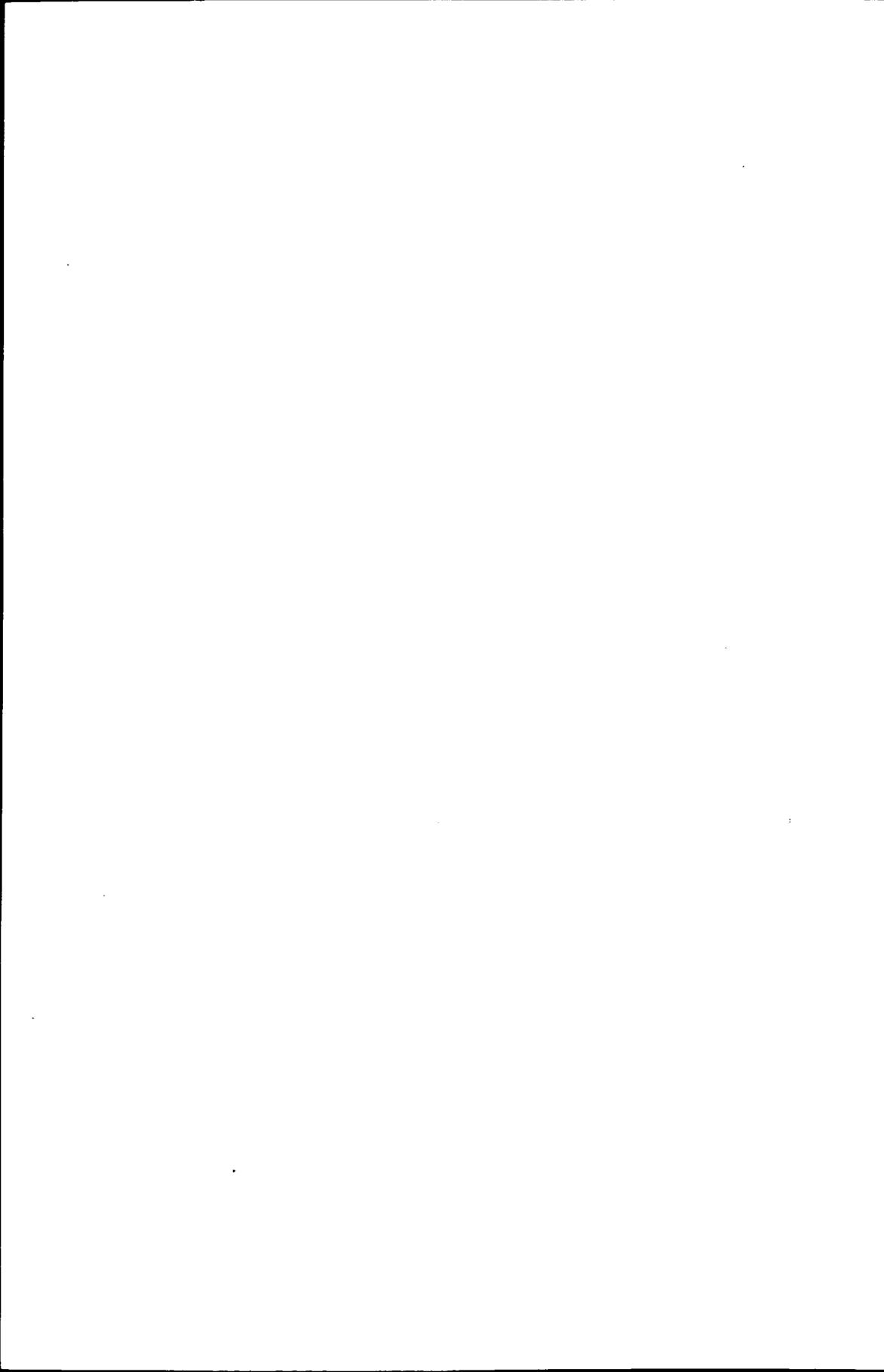
VI — Développer l'évaluation des politiques publiques en France : cinq propositions

VI — 1	Rendre les politiques publiques évaluables	162
VI — 2	Organiser les compétences	163
VI — 3	Renforcer l'évaluation des politiques publiques au sein de l'administration	164
VI — 4	Accroître la capacité d'évaluation externe à l'administration et favoriser sa diffusion dans la société	166
VI — 5	Identifier les thèmes prioritaires	167

Glossaire	169
------------------	-----

Annexes	175
----------------	-----

Introduction



Introduction

L'évaluation des politiques publiques, c'est-à-dire, selon une définition sommaire, l'analyse *ex post* de leurs effets réels, fait l'objet d'une attention croissante dans notre pays.

En premier lieu, au plan des doctrines ou dans les débats d'idées. Les 1^{re} et 2^e lois de Plan soulignent la nécessité de développer « les moyens d'évaluation de la mise en œuvre et des résultats effectifs du 9^e Plan » par les administrations concernées et, également, les partenaires sociaux (encadré n° 1). Dans un passé proche, plusieurs colloques réunissant décideurs et experts ont été consacrés à ce thème. Il en est ainsi du colloque des 15 et 16 décembre 1983¹ organisé à l'École nationale d'administration à l'initiative de la Direction de la prévision et en collaboration avec l'École des hautes études commerciales², ou, encore, de celui portant sur le thème plus précis de l'évaluation des politiques sociales, des 27 et 28 septembre 1984, animé conjointement par le Commissariat général du plan et l'Association d'économie sociale. Divers groupes de travail administratifs se sont vu confier des missions d'expertise sur les modalités du développement de l'évaluation des politiques publiques, dans tel ou tel secteur. On mentionnera le groupe du Commissariat général du plan sur l'évaluation des effets locaux des politiques sociales transversales³ ou encore celui mis en place par le ministère de la Recherche et de la Technologie, qui avait pour vocation de définir les procédures d'évaluation des grands programmes de recherche (procédures prévues dans la 2^e loi — programme sur la recherche et le développement technologique)⁴.

Au plan institutionnel, plusieurs organisations se réclamant de l'évaluation sont d'ores et déjà en place. Aux précurseurs tels que le Centre de prospective et d'évaluation commun au ministère du Redéploiement industriel et du Commerce extérieur et au ministère de la Recherche et de la Technologie, se sont joints divers nouveaux venus. Mentionnons, ainsi, l'Office parlementaire d'évaluation des choix scientifiques et technologiques⁵, destiné à réaliser des travaux d'évaluation pour l'Assemblée nationale et le Sénat, ou le Comité national d'évaluation des

1. Les communications faites à ce colloque et un résumé des débats figurent dans l'ouvrage édité par J.P. Nioche et R. Poinsard *L'évaluation des politiques publiques* (1984, Economica). Cet ouvrage est désigné ci-après par le sigle EPP.

2. Ce colloque a également bénéficié du concours de l'École nationale des ponts et chaussées et de la participation du Secrétariat d'État chargé de la Fonction publique et des Réformes administratives.

3. Voir le *Rapport sur l'évaluation des politiques sociales transversales : une méthodologie d'évaluation de leurs effets locaux* (décembre 1985, Commissariat général du plan ; président du groupe : J.-C. Ray ; rapporteurs : J.-M. Dupuis, J.-P. Tricart).

4. Voir le *Rapport sur l'évaluation des programmes* (novembre 1985, ministère de la Recherche et de la Technologie ; Service de la programmation et de la recherche ; président du groupe : J.-M. Martin).

5. On trouvera une analyse des fonctions de cet organisme, dont la création a été partiellement inspirée par l'exemple américain de l'Office of technology assessment, dans les communications de MM. Quermonne et Rouban, et Dupas et Klein, sur le thème de l'évaluation et du contrôle politique (in EPP).

établissements supérieurs d'enseignement et de recherche⁶. Des organismes d'étude et de recherche réorientent leurs activités vers le domaine de l'évaluation, tel le CREDOC qui entend faire de l'évaluation des politiques sociales un de ses axes majeurs de travail.

Cette double dynamique de pensée et d'organisation s'est accompagnée d'un accroissement effectif des études d'évaluation⁷. On peut mentionner plus particulièrement la réalisation de plusieurs études interministérielles⁸ et l'inclusion de plus en plus fréquente de procédures d'évaluation dans la définition même des programmes d'action publique. Il en a été ainsi pour le plan textile ou les TUC⁹.

Notons enfin que le terme même d'évaluation connaît une faveur sémantique indéniable : les projets de liaison fixe transmanche ont été jugés par un groupe d'évaluation, il a été décidé de créer une fondation pour l'évaluation des techniques et pratiques médicales¹⁰...

*
**

6. Ce comité, présidé par M. Laurent Schwartz, a été mis en place dans le courant du deuxième trimestre de l'année 1985.

7. On reviendra sur ce point dans la partie V du rapport. Le bilan établi par J.-P. Nioche et R. Poinard sur *L'évaluation des politiques publiques en France* (in EPP) fait apparaître un développement net des études d'évaluation à partir de la fin des années 1970 (de quelques unités par an à quelques dizaines).

8. Diverses études ont été lancées au cours du 2^e semestre 1984 suivant la procédure des « études prioritaires RCB ». Elles portent sur les sujets ci-après :

— la planification hospitalière

(ministères de l'Économie, des Finances et du Budget et des Affaires sociales et de la Solidarité nationale ; chef de mission : professeur Jean-Pierre Etienne) ;

— les aides à l'emploi

(ministères de l'Économie, des Finances et du Budget, du Plan et de l'Aménagement du territoire, du Redéploiement industriel et du Commerce extérieur, de l'Emploi et de la Formation professionnelle ; chef de mission : M. Pierre Chavance) ;

— l'efficacité des interventions publiques en faveur de l'élevage bovin (ministères de l'Économie, des Finances et du Budget et de l'Agriculture ; chef de mission : M. Michel Colonna).

Ce programme résulte d'un recentrage des activités d'étude conduites au titre de la RCB sur le thème de l'évaluation des politiques publiques.

On peut à ce titre citer également la mission interministérielle sur les exportations françaises de bâtiment et de travaux publics (ministères de l'Économie des Finances et du Budget, du Redéploiement industriel et du Commerce extérieur, de l'Urbanisme, du Logement et des Transports ; chef de mission : M. Gilbert Dreyfus, puis M. Jacques Bourdillon).

Un objectif de cette mission était d'évaluer l'efficacité des interventions de l'État dans ce domaine (aides à l'exportation).

9. On reviendra sur ces exemples dans les parties II et III.

10. La création de cette fondation, en décembre 1985, fait suite au rapport établi par le professeur Emile Papiernik à la demande du secrétaire d'État chargé de la Santé. On évoque dans la partie I, le système des « conférences-consensus » sur lequel s'appuie le fonctionnement de cette fondation.

L'évaluation des politiques publiques : les orientations du 9^e plan

« Au-delà du seul contrôle administratif d'exécution, il est nécessaire que soient développés les moyens d'évaluation de la mise en œuvre et des résultats effectifs du plan, par le Commissariat général du plan et par les ministères.

L'évaluation est nécessairement contradictoire et fondée sur le débat ; au-delà de l'évaluation par l'exécutif de sa propre action, le 9^e plan encouragera le développement de moyens d'expertise auprès des partenaires sociaux.

De même, les échelons régionaux de la planification décentralisée devraient créer leurs propres moyens d'évaluation, en s'appuyant notamment sur les moyens universitaires disponibles.

Enfin, les études et recherches sur les méthodes d'évaluation elles-mêmes sont trop peu développées en France ; il conviendra donc de poursuivre les travaux appliqués et méthodologiques en matière d'évaluation dans les services d'études et les centres de recherche publics, et d'assurer une coordination de ces travaux d'évaluation au sein d'une instance adaptée. Des actions incitatives devront également encourager ces travaux dans les centres universitaires et les centres d'études proches des partenaires sociaux » (1^e loi de Plan — Loi n° 83-645 du 13 juillet 1983 — Rapport annexe ; JO du 17 juillet 1983, p. 96).

« Enfin, de manière non directement liée au suivi d'un objectif particulier du 9^e plan mais en vue de disposer de moyens d'appréciation des conditions de son exécution et d'évaluation de ses résultats, des progrès seront accomplis dans les instruments de mise en œuvre de la planification, suivant plusieurs directions...

L'analyse des résultats des politiques engagées pendant le 9^e plan, résultats toujours complexes et d'une interprétation délicate, fera appel au développement des méthodes d'évaluation des politiques publiques. Il importe que les différents départements ministériels prennent des initiatives en ce sens, de façon que l'introduction de ces méthodes s'étende progressivement à l'ensemble de l'administration. Ces études devraient servir à réorienter en tant que de besoin les ressources budgétaires dont l'allocation optimale est nécessaire dans le contexte des finances publiques du 9^e plan.

Le Commissariat général du plan incitera les échelons régionaux de planification à créer leurs propres instruments d'évaluation en s'appuyant notamment sur les directions régionales de l'INSEE et les centres de recherche universitaire. » (2^e loi de Plan — Loi n° 83-1180 du 24 décembre 1983 — Rapport annexe ; JO du 30 décembre 1983, p. 53 et 54)

Dans cette dynamique, la référence au cas des États-Unis, grands innovateurs en la matière, a pu constituer un aiguillon appréciable.

Un examen rapide des discussions, l'observation de certains courants d'opinion reflétés, par exemple, dans le succès d'ouvrages ou d'émissions télévisuelles¹¹, montrent que, bien au-delà d'une imitation du modèle américain, la démarche évaluative répond à diverses attentes, parfois fortes, des acteurs sociaux. Elle se trouve de ce fait en accord avec plusieurs thèmes actuels de débat public. Là encore, on ne mentionnera ces attaches que de façon impressionniste.

Tout d'abord, l'évaluation peut apparaître, notamment à l'administration, comme un outil utile pour gérer des deniers publics devenus rares. Permettant de discriminer entre les « bonnes » et les « mauvaises » politiques, elle permettrait d'accroître l'efficacité (jugement sur les résultats), et non simplement l'efficience (jugement sur la mise en œuvre des moyens) de l'action de l'État.

Plus globalement, l'évaluation autoriserait une approche plus lucide, moins passionnée, de la délimitation des rôles de l'État par l'identification précise, « technique », des domaines où son intervention apparaît économiquement et socialement justifiée et de ceux où, au contraire, l'effet de ses actions est aléatoire, voire pernicieux.

L'évaluation des politiques publiques serait donc un instrument de changement. En cela, elle favoriserait cette modernisation, réclamée par tous, de l'État et de la société civile. Impliquant comparaison (aux objectifs, à un groupe témoin, à la réalité étrangère) et ouverture (appréciation des effets, donc de l'impact sur « l'extérieur »), elle constituerait un substitut à la sanction, concurrentielle, du marché. Instrument de sélection, elle préciserait également la responsabilité des décideurs gouvernementaux et administratifs vis-à-vis des politiques qu'ils définissent ou exécutent. L'évaluation permettrait enfin une adaptation plus continue et plus rapide des actions de l'État. Elle constituerait un frein à la sédimentation de dispositifs multiples, concurrents ou incohérents, et à la constitution, cumulative et peu réversible, de systèmes d'intervention opaques dont l'efficacité réelle est difficile à apprécier (cas fréquemment cité des aides publiques aux entreprises).

Au total, par sa nature pragmatique, par sa subordination aux faits, l'évaluation des politiques publiques peut apparaître, à beaucoup, comme relevant du « parler vrai ». Antidote contre les idées reçues, la langue de bois et la politique gadget, elle contribuerait à éclairer le dialogue social en dégagant le domaine des « réalités » de celui des « valeurs ».

*
**

11. La France n'en n'est certes pas encore au stade des États-Unis, où, en 1984, un journal comme le *Washington Post* a publié une centaine d'articles traitant de cas d'évaluation des politiques publiques (voir partie V du rapport). Le succès remporté par des ouvrages comme ceux de F. de Closets *Toujours plus, Tous ensemble*, ou des émissions telles que *L'Enjeu*, qui ont pour trait commun de présenter des enquêtes factuelles sur des thèmes souvent connexes à l'évaluation de politiques (ou d'organisations) publiques, est toutefois symptomatique.

Ce développement d'initiatives, cette multiplicité d'attentes peuvent faire penser que le départ sans retour vers les « mille fleurs » de l'évaluation est chose faite. Un examen rapide montre que rien n'est moins sûr.

Tout d'abord, la bonne volonté, ou même l'enthousiasme, ne sauraient suffire pour le développement et la consolidation de l'évaluation des politiques publiques. Le mouvement évaluatif, comme tout autre, traîne avec lui un halo, qui recouvre un grand flou conceptuel (lié notamment, et banalement, aux acceptions diverses du terme évaluation), des doctrines partiellement contradictoires, et une méthodologie approximative. Or l'évaluation de politiques publiques peut conduire à des remises en cause. En situation conflictuelle, les risques de déconsidération et, à terme, de dilution, sont forts si l'ancrage du mouvement évaluatif n'est que de sensibilité sans être technique et institutionnel. L'évaluation, pour perdurer, ne saurait demeurer un exercice d'honnête homme ou de militant. Trop d'échecs, par imprécision ou imprudence, peuvent paralyser son développement en donnant un alibi aux partisans du *statu quo*.

Ce risque est d'autant plus aigu que les attentes vis-à-vis de l'évaluation ne sont pas exemptes d'ambiguïtés.

La première, fondamentale, touche à la perception même du rôle de la puissance publique. On voit coexister, au sein de la société française et des individus qui la composent, des attitudes ponctuellement très critiques, sinon virulentes, vis-à-vis des politiques publiques ou des fonctionnaires qui les appliquent, et, simultanément, des réflexes de religiosité à l'égard de l'État dans son ensemble. Le mythe de l'État recours omniscient demeure plus répandu dans notre société qu'on veut bien le dire parfois. De semblables attitudes de révérence se retrouvent chez les experts ou les responsables de politiques sectorielles, indépendamment de tout esprit corporatiste. S'ils concèdent volontiers, que sur un grand nombre d'actions ponctuelles (la plupart ?), l'appréciation d'efficacité est des plus maigres, ils affirment simultanément que le bien-fondé de la politique globale ne doit pas être remis en cause : les malheurs du monde ne sauraient dissimuler que la grâce divine est à l'œuvre.

Cette situation peut être reliée à une tradition française qui privilégie les débats d'idées fondés sur le maniement de concepts ou de systèmes globaux au détriment d'approches plus factuelles. En matière de politiques publiques, cette tradition est susceptible de conduire au détournement de notions fondamentales (indépendance nationale, qualité du service public, équité) comme paravents à la perpétuation d'avantages particuliers ou au maintien de fonds de commerce administratifs.

Enfin, la faveur affichée pour l'évaluation ne doit pas faire totalement illusion. Il en est de l'évaluation comme de la concurrence : chacun y est favorable mais, de préférence, pour son voisin.

*
**

La mission confiée à notre groupe s'inscrit donc, pour partie, dans un contexte porteur. Il existe désormais une bonne réceptivité au concept d'évaluation, beaucoup plus large en tout cas qu'il y a seulement deux ou trois ans. Mais il convient également de reconnaître les chausse-trapes qui demeurent.

Dans cette perspective, compte tenu du mandat du groupe et de ses travaux, trois objectifs sont assignés au présent rapport : clarifier, informer, proposer.

Clarifier. Il a été dit que le concept même d'évaluation est utilisé de manière ambiguë. Il peut, selon les interlocuteurs, recouvrir des notions diverses : *ex ante* ou *ex post*, de politiques ou d'organisations... En outre, la pratique de l'évaluation se situe, au plan technique et conceptuel, à la confluence de plusieurs domaines : analyse des politiques, modélisation, expérimentation... Par ailleurs, il existe des genres voisins (contrôles administratifs, de gestion), vis-à-vis desquels la distinction peut apparaître plus de degré que de nature. D'où, globalement, un risque sérieux de dérive vers des conceptions très globales, à un niveau de généralité tel que toute spécificité disparaît.

Deuxième objectif : informer.

Cet objectif concerne notamment les plans de recherche et les techniques de l'évaluation. Il convient d'être clair sur ce point. Les travaux du groupe n'avaient pas pour ambition d'innover, d'aboutir à une nouvelle méthode miracle d'évaluation des politiques publiques. Il existe de bonnes « boîtes à outils », aux qualités éprouvées, ayant fait l'objet de nombreux exposés et discussions, notamment dans la littérature anglo-saxonne. Le rapport ne saurait, dans ce domaine des méthodes, apporter du neuf aux spécialistes.

Il se doit d'éviter la tentation d'une originalité qui demeurerait strictement nominaliste. Il conserve par contre la fonction importante d'informer, de manière accessible mais précise, un public élargi.

En matière d'information, le rapport doit également tirer parti des nombreux exemples présentés lors des travaux du groupe ou décrits dans des publications diverses. Il convient, dans le domaine des applications, de ne pas en rester à l'émiettement apparent des évaluations existantes, mais de remettre en cause les arguments, souvent évoqués et largement spécieux, de complexité et de spécificité des politiques publiques qui feraient de chacune d'elles un cas d'espèce. Obtenir quelques enseignements généraux est à la fois utile et possible.

Enfin, le rapport se conclut par des **propositions d'action**. Elles concernent à la fois la formulation des politiques publiques, le développement des compétences nécessaires, le rôle de l'administration, la diffusion de l'évaluation dans la société. Leur mise en œuvre ne saurait être fixée en détail dans ce rapport. Elle appelle une action soutenue, notamment des Pouvoirs publics. Pour les membres du groupe, les orientations proposées sont les garants d'un développement solide et pluraliste de l'évaluation des politiques publiques, adapté à l'environnement national.

I

L'évaluation des politiques publiques : définition et frontières



I

L'évaluation des politiques publiques : définition et frontières

I - 1 Les risques de l'imprécision

Nous avons dit que le mouvement, qui se dessine en France en faveur de l'évaluation des politiques publiques, est menacé par certains risques de dilution. Faute d'assise méthodologique et déontologique, l'excès des ambitions comme le vague des références peuvent conduire à l'échec au même titre que les obstacles institutionnels.

Il est donc essentiel de dissiper d'emblée de possibles malentendus.

Le premier tient à l'ambiguïté du terme évaluation, utilisé dans des acceptions très diverses. Nous avons dit que plusieurs instances se réclament du concept, mais plus au sens d'analyse ou d'appréciation¹ que de véritable mesure *ex post* et d'identification de causalités. On constate aussi un usage du terme assez indistinct pour les études « *ex ante* » et « *ex post* ». Ces deux domaines présentent des frontières communes. Ils appellent pour partie des techniques semblables. Toutefois, il importe, pour dégager des recommandations efficaces, de les distinguer nettement au plan conceptuel.

Le deuxième risque de glissement tient au succès même du concept d'évaluation. On a mentionné les différentes forces qui lui confèrent actuellement un caractère porteur mais sans finalité très explicite. L'intérêt porté à l'évaluation comporte ainsi, dans notre pays, une composante de **mode** et contient le danger symétrique d'une dévalorisation rapide.

Troisième facteur de risque : l'évaluation des politiques publiques traite de **domaines « mous »**. La réduction nécessaire des phénomènes à une description susceptible d'analyse, en particulier quantitative, la définition explicite de plans de recherche impliquent une discipline qui n'est pas toujours suivie. Certes, il convient d'éviter les excès d'un positivisme mécaniste et naïf. On insistera ultérieurement sur l'importance d'une référence préalable à une « histoire » de la

1. Ou encore concernant plus la phase de formulation d'une politique, que la mise en évidence de ses résultats.

politique évaluée². Il demeure qu'à l'heure actuelle les écueils les plus sérieux en matière de méthodologie paraissent plus liés aux tentations de la logomachie³ qu'à celles du purisme technique.

Un autre danger tient à ce que l'évaluation des politiques publiques est un levier de remise en cause. Elle s'applique à des domaines où les conflits de pouvoir sont souvent forts, où les intérêts catégoriels peuvent secréter des idéologies protectrices et amener parfois des connivences, implicites ou explicites, entre administrations et lobbies professionnels. Faute de garde-fous, sympathie ou intérêt peuvent aisément conduire à la compromission. En situation conflictuelle, un risque connexe est celui d'un usage partisan de l'évaluation. Évaluations et contreévaluations, bilans et contrebilans, s'échangent entre camps adverses. S'il n'existe pas un minimum de normes admises par tous, une telle pluralité des évaluations, *a priori* souhaitable, jette le discrédit sur la démarche même, ressentie comme permettant de prouver « n'importe quoi »⁴.

Une dernière série de dangers tient aux ambitions, ou aux enthousiasmes, de certains partisans sincères de l'évaluation. Plus qu'une discipline d'analyse et de constat, elle est parfois conçue comme un instrument de consensus social, propre à favoriser le développement de pratiques participatives et la résolution des conflits, ou encore susceptible d'amener des formes nouvelles de rapports sociaux au travers d'expérimentations. Ces conceptions, qui présentent des aspects positifs, sont également porteuses de risques notables.

L'évaluation-consensus, c'est-à-dire avec la participation étroite des acteurs de la politique étudiée, est souvent préconisée dès lors qu'une organisation publique est concernée. On la justifie par diverses considérations. Certaines, qui paraissent légitimes, confèrent à cette association un rôle pour l'efficacité de l'évaluation. Ainsi, lorsqu'une organisation est en jeu, il peut être difficile de réaliser une évaluation contre elle, pour ainsi dire en force. En outre, la consultation « d'opérationnels » pendant la procédure d'évaluation permet d'éviter les naïvetés attachées à un point de vue trop exclusivement de Sirius. Au-delà de ces remarques de bon sens, il est par contre beaucoup plus contestable de s'engager dans une voie participative extrême, où l'évaluation se réduit à une thérapie de groupe, propre à

2. D.T. Campbell insiste à juste titre sur l'importance de la « narrative history » (voir D.T. Campbell (1975) *Assessing the impact of planned social change* in *Social Research and Public Policies*, G.M. Lyons ed. The Public Affairs Center, Dartmouth College, Hanover, New Hampshire). Elle est d'autant plus forte que la politique concernée est « complexe » (dans son champ, ses effets, sa durée, etc.).

3. Pour s'en tenir au cas des États-Unis, où la tradition empirique est plus forte qu'en France, l'étude de J.A. Schneider, N.J. Stevens, L.G. Tornazky *Policy research and analysis : an empirical profile, 1975-1980* (Policy Sciences, Vol. 15 n° 2, décembre 1982), portant sur un échantillon de 181 articles publiés dans diverses revues (12 % du nombre total d'articles), fait ressortir qu'une minorité seulement (46 %) recourent à un plan de recherche explicite (et sont alors de style « empirique-quantitatif »). La majorité (54 %) relève, selon les auteurs, de la catégorie « discussion rhétorique ».

4. Par exemple, on peut estimer que ces risques ont été sous-estimés dans les évaluations faites de la loi sur le prix unique du livre. La confrontation d'évaluations très hétérogènes et aux méthodologies insuffisamment explicitées permet difficilement au citoyen-lecteur de se faire une opinion précise. Une liste non exhaustive comprendrait : — une évaluation de style « consensus » faite par l'Observatoire du prix des livres qui associe des professionnels tels que les libraires et les éditeurs ;

— P. Messerlin — *Le prix du livre, analyse économique de la loi Lang*, Institut La Boétie, 1984 ;

— C. Fauvelais et J.Y. Blain — *Le prix unique pour le livre*, Editions de l'Institut économique de Paris, 1983 ;

— B. Pingaud — *Le livre à son prix*, ministère de la Culture, 1983.

résoudre certaines difficultés internes à l'organisation. Une telle démarche présente des risques de collusion qui vont à l'encontre des ambitions même de l'évaluation des politiques publiques telles qu'elle est comprise dans ce rapport⁵.

On reviendra ultérieurement sur cette question importante de l'association des « évalués » à l'évaluation (paragraphe III-5).

Plus généralement, la conception de l'évaluation comme substitut à la décision est à rejeter. Elle a été menée à un point extrême par les tenants de « la société expérimentale »⁶. D'après ceux-ci, la réalisation planifiée d'expériences sur petits groupes est un moyen d'inventer sans frais la société de demain. On rappellera que l'histoire ne se met pas en éprouvette. L'évaluation ne peut se substituer au choix de valeurs qui constituent l'essence des choix de politiques : elle peut seulement, au mieux et à certaines conditions, permettre un avis relativement fiable sur les conséquences de ces choix. Une condition pour cela est qu'elle restreigne son champ. Il s'agit de fournir un élément d'information, parmi d'autres, à l'opinion et aux décideurs politiques sur des domaines suffisamment limités pour pouvoir faire l'objet d'une analyse de type scientifique. **La démarche scientifique de l'évaluation ne doit en aucun cas être confondue avec les illusions du « gouvernement scientifique ».** Tout au plus, peut-elle s'affirmer comme un point de passage utile, voire obligé, en termes de procédure.

Au total, le développement de l'évaluation des politiques publiques requiert de se référer à un **code méthodologique et déontologique comportant un minimum de règles**. Il ne s'agit pas d'enfermer les pratiques de l'évaluation dans un réseau d'obligations conformistes ni de décourager son développement par des exigences excessives de rigueur. Il importe par contre de reconnaître les facteurs potentiels d'échec les plus importants pour le développement de l'évaluation. A l'heure actuelle, et pour le cas français, ils paraissent se situer au moins autant du côté du flou et de l'inorganisé que de celui de la rigidité et de la sclérose.

Une certaine « institutionnalisation » de l'évaluation posséderait une dernière vertu : permettre un processus **d'accumulation**. Comme le note B. Eveno dans sa communication sur les rapports entre évaluation et contrôle administratif (in EPP), la vérification de cinquante hôpitaux ou universités suivant des approches largement spécifiques à chaque contrôleur conduira à des rapports qui ne seront ni fongibles ni additionnables. Cette remarque garde sa valeur pour les évaluations de politiques publiques, où l'objectif d'une généralisation des résultats est souvent primordial.

5. Ces observations sur l'évaluation-consensus ne signifient pas que des pratiques « associatives » sont dépourvues de valeurs, mais pour d'autres finalités ou à d'autres stades du « cycle politique ». Par exemple, les « conférences consensus » sur lesquelles doit s'appuyer le fonctionnement de la Fondation pour l'évaluation des techniques et des pratiques médicales, ont comme objectif de faire la synthèse d'évaluations partielles pour formuler des recommandations de politique. De même « l'évaluation » des DRIR (Directions Régionales de l'Industrie et de la Recherche) réalisée par le CPE (Centre de Prospective et d'Évaluation) a fait appel à une participation importante des acteurs concernés (voir encadré n° 9, partie II, p. 83). Il s'agit, là encore, plus de procédures de prise de conscience ou de formulation de politiques, extrêmement utiles pour la dynamique d'une organisation, que d'évaluation *ex post stricto sensu*.

6. Voir notamment D.J. Campbell *Methods for the experimenting society*. Working Paper, Department of psychology, Northwestern University Evanston. Selon cette conception, les hommes politiques sont les « créatifs-imaginatifs » qui suggèrent des voies nouvelles pour la société. La sélection de leurs propositions est faite au travers des pratiques froides des évaluateurs, notamment l'expérimentation.

I - 2 Définition et attributs de l'évaluation

La discussion précédente conduit donc à retenir une définition stricte de l'évaluation. Plusieurs définitions, proches de cette conception, ont déjà été proposées. Nous avancerons la suivante, qui en réalise une certaine synthèse.

Évaluer une politique publique, c'est reconnaître et mesurer ses effets propres.

Quelques commentaires sur les termes de cette définition :

— la reconnaissance, l'identification des effets propres de la politique évaluée indique clairement que le processus d'évaluation passe par une certaine **reconstruction** du « réel ». L'évaluation des politiques publiques ne peut, en général, se limiter à une démarche passive d'observation de données brutes ;

— le terme de **mesure** fait clairement référence à l'ambition quantitative de l'évaluation. Cette ambition ne constitue pas une exclusive à l'égard d'approches qualitatives, les études de cas par exemple. Ces approches sont, au contraire, un préalable souvent indispensable à la mise en évidence d'ordres de grandeur. Il demeure que l'évaluation d'une politique publique, pour être complète, doit aller au-delà d'une simple description qualitative chaque fois que les circonstances s'y prêtent, ce qui est presque toujours le cas ;

— cette mesure concerne les **effets** de la politique. On fait donc référence à un concept de **causalité**. Cette partie de la définition renvoie ainsi, pour partie, à une approche des politiques publiques en terme de théorie du changement (voir I-4) : les politiques publiques sont supposées être mises en œuvre pour produire certains effets sur un « milieu » extérieur. Dans cette perspective, l'évaluation d'une politique publique apparaît comme un **test d'hypothèse**.

— il est précisé enfin que les effets qui doivent être identifiés et mesurés sont les **effets propres** de la politique, non les effets attendus, objet de l'analyse *ex ante*. Ce terme suppose également que l'évaluateur est capable de discriminer, dans l'observation des résultats obtenus, ce qui revient à la politique proprement dite et ce qui revient à l'environnement, ce dernier concept étant pris dans un sens large.

La **définition choisie implique que les attributs de la démarche évaluative ainsi comprise sont, intrinsèquement, ceux de la démarche expérimentale, même si les techniques utilisées ne peuvent être toujours identiques, en raison de la spécificité des domaines d'application**⁷.

Ces attributs sont les suivants : **identification, mesure, confrontation, explication**. Nous allons les mettre en parallèle avec les termes de la définition précédente.

1) L'évaluation doit tout d'abord se fonder sur une **identification précise du système à évaluer**. Elle concerne l'objet de l'évaluation, c'est-à-dire « la » politique évaluée, les objectifs de cette politique, les facteurs d'environnement pertinents

7. Ce point souvent souligné, mais aussi souvent oublié, indique bien que le **concept d'évaluation proposé ne correspond pas à une mythification du « big science model »** (voir D.T. Campbell « Assessing etc. » op. cité). On soulignera ultérieurement que les méthodes dites expérimentales ne peuvent être systématiquement envisagées pour l'évaluation des politiques publiques (même si elles peuvent l'être en France beaucoup plus qu'actuellement !). Il demeure que la logique de la démarche expérimentale, sa discipline, nous paraissent, elles, valides dans tous les cas.

(nous gardons ce terme vague pour le moment). Une difficulté centrale de cette phase de l'évaluation est de mettre de l'ordre dans un ensemble de **données multidimensionnelles et interdépendantes**. Il est nécessaire à cette fin de procéder explicitement à des coupures, à des bornages du champ étudié. Faute de cette discipline, on aboutit à des analyses où tout est dans tout, exhaustives certes, mais faiblement conclusives.

Une **première délimitation** concerne l'**objet même de l'évaluation**, à savoir la **politique**. Il n'est pas toujours facile d'y procéder. L'évaluateur est rarement confronté au cas simple d'une mesure isolée et spécifique, les politiques à évaluer constituant le plus souvent des objets complexes⁸ mettant en œuvre des moyens à usages multiples. Une politique particulière peut ainsi être insérée dans un programme plus large, dont elle est un élément constitutif. Ou encore, elle se situe à la marge d'un système préexistant, et n'est qu'un dispositif parmi d'autres, gérés par la même organisation et de finalités similaires. Cette difficulté à « marginaliser » la politique évaluée par rapport à un système global se rencontre dans de nombreux domaines : éducation, santé, aides aux entreprises etc.

Une difficulté connexe pour délimiter l'objet de l'évaluation renvoie au thème de l'**analyse d'organisation**. Nombre de politiques publiques sont en effet exécutées au travers d'organisations spécifiques. Ces organisations peuvent être (relativement) simples dans le cas de politiques réglementaires ou de redistribution sociale. Elles sont, en général, plus complexes lorsqu'il s'agit d'une production de biens et services, en particulier non marchands (éducation, santé par exemple). Le succès d'une politique comme son échec peuvent alors tenir non seulement au réalisme des ambitions initiales, au caractère favorable de l'environnement, mais aussi aux performances de l'organisation support. Dans certains cas, l'évaluation d'une politique impose donc une analyse d'organisation assez fine (mais ne saurait s'y réduire).

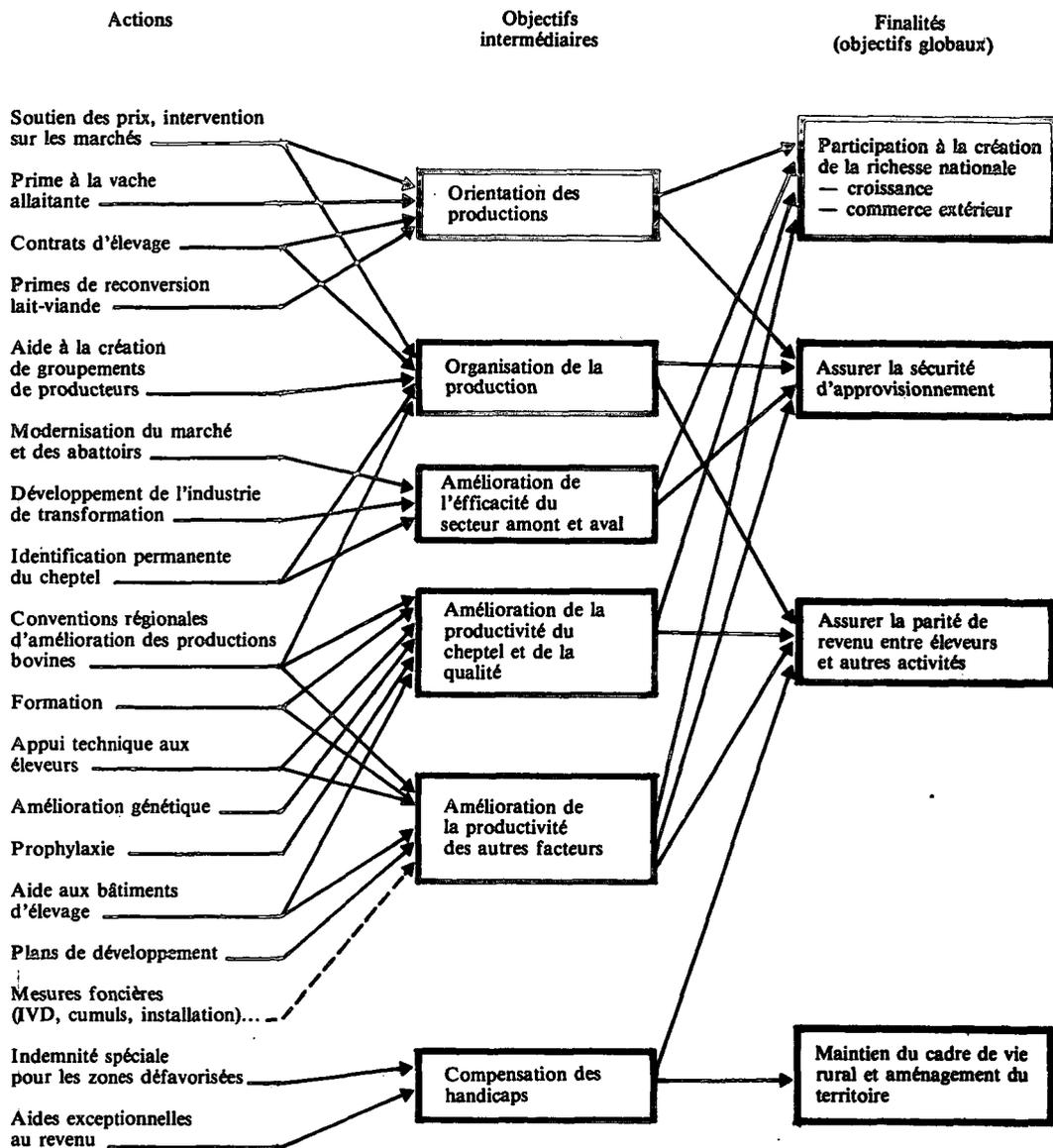
Une fois reconnu « l'objet » de l'évaluation, il convient d'identifier les **objectifs** poursuivis par la politique. Nous reviendrons sur ce thème ultérieurement, en distinguant entre **objectifs finaux**, souvent qualitatifs et généraux, et **objectifs intermédiaires** (ou opérationnels) traduisibles en indicateurs chiffrés. Si des objectifs finaux sont en général formulés lors de la mise en place d'une politique, l'affichage d'objectifs intermédiaires est très loin d'être systématique. Il peut donc revenir à l'évaluateur de « traduire » les objectifs finaux en objectifs intermédiaires, tâche d'autant plus ardue que les premiers sont très globaux⁹. A titre d'illustration, l'encadré n° 1 donne une description de la structure d'objectifs retenue dans l'étude interministérielle sur les interventions publiques en matière d'élevage bovin.

8. Le terme de « messy » (embrouillé) est fréquemment utilisé dans la littérature anglo-saxonne.

9. On mentionnera ainsi la référence faite à l'objectif d'« indépendance nationale » dans un grand nombre de politiques sectorielles (agriculture et industrie en particulier). Un tel objectif n'est pas facile à qualifier pour une économie de plus en plus ouverte et ayant largement dépassé le stade de l'autosubsistance.

Encadré n° 1

L'étude sur les interventions publiques en matière d'élevage bovin : la structure des objectifs¹⁰



10. Communication de C. Le Pen au groupe de travail (source : Mission d'analyse et d'évaluation du ministère de l'Agriculture).

Dernier stade de la phase d'identification : reconnaître les « facteurs d'environnement » pertinents. Ce problème est analogue à celui évoqué au sujet de l'analyse d'organisation et se résout également à partir d'une certaine « analyse de système », visant à préciser les variables non contrôlées par le décideur public et susceptibles de peser sur les effets de la politique (exemple : facteurs socio-culturels en matière d'éducation, groupes de pression etc.). Des opérations de coupure et de bornage sont, là aussi, nécessaires. Il revient à l'évaluateur de distinguer parmi les interrelations celles qui lui paraissent secondaires ou majeures. Cette opération appelle un jugement de sa part qui ne peut être éludé. Il est parfois difficile d'atteindre un choix totalement incontestable. Par contre, la **reproductibilité de l'évaluation**, au sens d'une possibilité de vérification par d'autres, peut être garantie dès lors que ce choix est affiché explicitement.

Au total, cette discussion de la phase d'identification fait ressortir qu'une évaluation ne peut se faire de manière totalement mécaniste (sauf mesure extrêmement ciblée : action des sels fluorés sur la carie dentaire par exemple — voir partie II). **Une phase de reconnaissance qualitative, consistant, comme on l'a dit, à écrire une « histoire » de la politique est en général un préalable nécessaire, d'autant plus que la politique évaluée est complexe (au moins en apparence)¹¹.**

Cette phase narrative ne saurait toutefois constituer une évaluation complète. Certains auteurs limitent parfois le champ d'étude souhaitable à la seule analyse des procédures décisionnelles ou du fonctionnement des organisations, à la « compréhension » des politiques sans aller jusqu'à l'évaluation de leurs effets. Ils nous semblent encourager involontairement des attitudes conservatrices ou largement contemplatives.

2) La volonté de **mesure** est la deuxième caractéristique importante d'une démarche d'évaluation. Elle la distingue d'approches strictement qualitatives.

Comme on l'a dit, il ne s'agit pas de mythifier le chiffre. Par contre, l'ambition même de l'évaluation conduit à recourir à des indicateurs quantitatifs permettant d'apprécier l'ordre de grandeur des effets propres. Cet objectif est plus ou moins facile à atteindre. Il l'est d'autant plus que la politique est simple, clairement finalisée, et comporte des composantes matérielles identifiables. Des problèmes délicats surgissent lorsque la politique a des finalités globales et multiples et fait intervenir des caractéristiques largement immatérielles. Tel est le cas, par exemple, pour l'évaluation des effets locaux des politiques sociales transversales, qui a fait l'objet du groupe de travail du plan présidé par J.C. Ray (p. cité — encadré n° 2).

On reviendra sur ce problème de mesure à travers divers exemples. Mentionnons dès à présent que pour des politiques diffuses ou immatérielles, la parade habituelle est de se référer à une **batterie d'indicateurs**. Mentionnons également

11. Cette remarque s'applique, par exemple, à des politiques industrielles, étalées sur plusieurs années, ayant connu diverses péripéties, faisant intervenir des procédures de décision complexes à plusieurs acteurs (lobbies, administrations, etc.). Voir ainsi l'analyse de la politique sidérurgique faite par F.G. Padioleau (*Quand la France s'enferme*, 1981, PUF) ou celles du plan calcul et de l'électronucléaire présentées par F. Jenny (*L'évaluation des politiques publiques industrielles* in EPP).

que peut se poser un problème de perversion de ces indicateurs dès lors qu'ils sont affichés à l'avance¹².

3) La confrontation est une troisième caractéristique importante de la démarche évaluative. Parler des « effets » d'une politique signifie que l'on est capable de répondre à la question suivante : que se serait-il passé si la politique évaluée n'avait pas été mise en œuvre ? Cette exigence peut être satisfaite de manière plus ou moins complète comme on le verra ci-après.

Encadré n° 2

L'évaluation des effets locaux de politiques sociales transversales : la recherche d'indicateurs

Le Commissariat général du plan a constitué en janvier 1985 un groupe de travail ayant pour mission de contribuer à la définition d'un dispositif d'évaluation des politiques sociales transversales. Ce groupe présidé par le professeur Jean-Claude Ray, de l'université de Nancy II, a remis ses conclusions en décembre 1985.

L'évaluation des politiques sociales transversales soulève, de manière générale, les difficultés mentionnées et analysées ici, mais avec une acuité particulière : diversité des orientations et des modalités, importance des conditions locales, multiplicité des objectifs, variété des effets dans leur dimension temporelle et spatiale, implication de nombreux acteurs etc.

Le rapport propose, à la fois, des principes généraux pour l'évaluation (repérage des objectifs, recensement des moyens mis en œuvre, identification et mesure des effets, appréciation de résultat) et un « guide du bon usage des indicateurs » par grand domaine concerné (démographie et famille, logement et cadre de vie, éducation etc.). La source, le cadre territorial, la périodicité, les objectifs servis par l'indicateur et les précautions d'emploi sont précisés pour chaque indicateur proposé. Ainsi, en matière d'éducation (évaluation de la politique des zones d'éducation prioritaires) figurent des listes d'indicateurs de parcours scolaire (taux de redoublement, retards, orientations à divers niveaux de scolarité...) et d'offre scolaire (nombre d'élèves par classe, âge et formation des enseignants, distance domicile établissement, demandes de dérogation à la sectorisation...). La multiplicité et la variété des indicateurs proposés, leur niveau local, constituent une réponse aux difficultés spécifiques rappelées ci-dessus.

12. Un exemple classique est celui de la politique anti-crime du président Nixon (« crackdown on crime »). Plusieurs auteurs ont souligné que le résultat le plus tangible en a été la perversion des indicateurs manipulés « à la baisse » par les administrations locales concernées, pour faire apparaître des progrès (non enregistrement de certains délits, enregistrements dans des catégories de gravité inférieure etc.).

Il demeure que toute démarche d'évaluation cherche à **confronter l'état observé (« avec » la politique) à un autre état** : sans la politique, avant la politique, dans un autre pays etc. Cette exigence de **confrontation ouverte** distingue l'évaluation des politiques publiques d'autres approches, comme les contrôles administratifs ou de gestion (voir II-4). Elle marque sa parenté méthodologique avec la démarche expérimentale¹³.

4) Dernier attribut de l'évaluation des politiques publiques : la recherche d'une **explication** des effets observés. Cette recherche, qui répond à des interrogations de causalité, distingue également l'évaluation de politiques de procédures de contrôle. Dans ce dernier cas, l'objectif d'explication n'est pas absent mais garde un caractère d'exception : on « ira voir » seulement si le résultat n'est pas conforme à la norme initiale. Pour l'évaluateur, l'identité entre effets réels et effets attendus ne révèle pas *a priori* l'efficacité d'une politique, compte tenu, en particulier, du jeu des facteurs d'environnement.

Le niveau comme la fiabilité de l'explication obtenue peuvent toutefois être variables. Quelques mots sur ce point, discuté plus en détail avec la présentation des méthodes (partie II du rapport). On utilise pour ce faire des notions de **validité** traditionnelles dans la littérature sur l'évaluation¹⁴. Ces notions, qui renvoient à divers degrés d'analyse explicative, vont être rapidement précisées.

Le premier degré d'explication concerne la seule réponse à cette question fondamentale : y a-t-il un lien, ou non, entre la politique publique mise en œuvre et l'évolution observée des variables indicatrices de ses résultats potentiels ? Une réponse peut être obtenue si l'on est capable de construire le dispositif d'observation représenté dans le schéma 1.

	La politique est mise en œuvre	La politique n'est pas mise en œuvre, toutes choses égales par ailleurs.
État observé	E ₁	E ₂

Schéma 1

Un dispositif idéal d'observation pour l'évaluation

La difficulté tient évidemment au caractère virtuel de l'état E₂. On peut essayer de le construire par le biais d'une expérimentation véritable, avec d'ailleurs des pièges possibles, ou d'en atteindre une certaine approximation (techniques avant-après

13. Voir D. Schwartz *Les techniques d'évaluation en épidémiologie* (in EPP).

14. Il existe diverses présentations des concepts de validité, avec des variations suivant les auteurs. Nous adoptons ici un parti de simplicité en retenant une présentation concise qui distingue seulement trois types de validité : interne, conceptuelle, externe (voir par exemple R. Angelmar *Les méthodes de l'évaluation* in EPP). Les autres présentations consistent généralement en variantes, éventuellement désagrégées, de ce schéma central. Par exemple J.P. Cook et D.T. Campbell (*The design and conduct of quasi-experiments in field settings* (1976) in M.D. Dunnette ed. *Handbook of industrial and organizational psychology*. Rand Mac Nally, Chicago) introduisent un concept de validité statistique comme sous-catégorie de la validité interne etc.

etc.), ce qui soulève d'autres questions (on reviendra sur ces divers points dans la partie II). Notons également le caractère très exigeant, particulièrement pour l'expérimentation vraie, de la condition « toutes choses égales par ailleurs ». Quoi qu'il en soit, même en l'absence de ce dispositif idéal, l'évaluateur se doit de donner une réponse à cette première question de l'existence d'une relation de cause à effet entre introduction de la politique et évolutions observées. Cette question est à l'origine de la démarche même d'évaluation. Selon la terminologie consacrée, la fiabilité de la réponse reflète le degré de validité interne de l'évaluation¹⁵.

Ce premier degré d'explication sur la politique publique en tant que cause peut être difficile à atteindre. Il correspond pourtant à une analyse explicative qui demeure limitée. Constaté par exemple que la politique a réussi, n'implique pas nécessairement que l'on sache pourquoi, par quels mécanismes¹⁶. La réponse obtenue risque donc d'être insuffisante dès lors qu'on veut en généraliser la portée. Il peut s'agir tout d'abord, pour un même groupe cible, d'apprécier l'impact d'une modification de la politique évaluée ou la nature de résultats de niveau différent (par exemple le passage des effets à court terme aux effets à moyen terme). Suivant la terminologie consacrée, le caractère licite de telles généralisations, à groupe cible donné, renvoie à la validité conceptuelle de l'évaluation. D'autres interrogations concernent l'extension des conclusions sur l'efficacité, ou l'inefficacité, de la politique évaluée à d'autres cibles ou d'autres contextes. Prenons un exemple. On met en évidence qu'une politique de crédit d'impôt-recherche appliquée à un secteur particulier a eu un effet très positif sur l'innovation : ce résultat plaide-t-il, ou non, pour une extension de la mesure à d'autres secteurs ? La possibilité de répondre renvoie à la validité externe de l'évaluation.

Quels que soient les concepts retenus¹⁷, on voit intuitivement que l'objectif d'une généralisation des conclusions appelle des procédures particulières. En anticipant quelque peu, deux thèmes importants ressortent. Le premier est celui d'une « augmentation de la variance », selon la terminologie statistique consacrée : la modification des conditions contrôlées d'expérience ou la multiplication des observations permet d'enrichir l'information au-delà d'un seul constat ponctuel. Le second est celui de la construction de « modèles explicatifs ». Ces modèles, qui postulent des relations causales, testées systématiquement, peuvent permettre d'identifier les mécanismes à l'œuvre et de simuler les résultats de la politique dans des contextes divers.

15. Des listes de « pièges » limitant la validité interne d'une évaluation figurent dans plusieurs des références mentionnées (R. Angelmar, D.T. Campbell « Assessing... », T.J. Cook, et D.T. Campbell). On y reviendra dans la partie II.

16. Cette situation est caractéristique d'un grand nombre d'évaluations thérapeutiques. On peut parfaitement prouver l'efficacité d'un médicament sans être capable de reconstituer l'ensemble des mécanismes bio-chimiques à l'œuvre.

17. Dont il ne faut pas dissimuler le caractère partiellement arbitraire qui autorise les variations entre auteurs mentionnés plus haut.

I - 3 Les genres voisins de l'évaluation : aide à la décision, contrôle de gestion, contrôles administratifs

La définition stricte de l'évaluation proposée ci-dessus permet de mieux marquer sa spécificité et ses complémentarités par rapport à d'autres approches. Nous traiterons successivement des **rappports entre évaluation et aide à la décision, contrôle de gestion, contrôles administratifs.**

1) Bien que naturellement orientée vers la production de connaissances utiles à l'action, ou de propositions d'action, l'évaluation opère d'abord un retour en arrière. L'étude *a priori* d'aide à la décision, est, par définition, tournée vers le futur.

Il n'y a pas là qu'une nuance chronologique par rapport au moment de la décision et au déroulement de la politique considérée. L'étude *a priori* prend nécessairement un caractère **normatif**, ce qui n'est pas le cas de l'évaluation. Vis-à-vis de l'action, on retrouve cette différence de finalité. L'analyse *a priori* procède, fondamentalement, d'une démarche d'**optimisation** (plus ou moins complète) et tente d'éclairer le choix de « la meilleure » politique publique. L'évaluation fournit des enseignements de nature « **incrémentaliste** » pour l'action, c'est-à-dire autorisant de l'ajuster par pas successifs : les conclusions d'une évaluation peuvent permettre que les politiques à venir soient « **moins mauvaises** » que les précédentes¹⁸.

Ceci étant, les deux démarches présentent des liens évidents. Au plan des techniques utilisées tout d'abord. Elles sont fréquemment similaires. Par exemple un modèle explicatif (voir partie II) peut être aussi bien utilisé pour éclairer *a priori* une décision (simulations *ex ante*, optimisation) que pour évaluer une politique (simulations *ex post* prenant en compte l'environnement observé). De même, la démarche expérimentale, caractéristique de l'évaluation, peut être utilisée dans une perspective d'éclairage *a priori*, pour savoir si telle politique mérite d'être appliquée de manière générale¹⁹.

En outre, l'analyse *a priori* et l'évaluation *ex post* sont des démarches fortement complémentaires. La réalisation de la première est un gage d'efficacité accrue pour la seconde. Les critères employés sont alors explicites, ainsi que le contrat liant les services chargés d'appliquer la politique. L'évaluation suit naturellement, comme vérification de l'atteinte des objectifs ou de la validité des calculs réalisés *ex ante*. Cette complémentarité ressort bien des travaux effectués sur le cas de la périnatalité²⁰ : à l'issue d'une étude RCB, un programme de 7 actions a été mis en place (en 1971) grâce à des mesures incitatives et réglementaires. Trois enquêtes

18. On retrouve ici la différence entre démarche d'optimisation et démarche heuristique (search theory) classique en théorie de la décision.

19. Un exemple caractéristique est fourni par l'expérience de l'impôt négatif du New-Jersey décrite dans la partie II. On notera que l'exigence d'explication, d'analyse des mécanismes est particulièrement forte dans l'analyse *a priori*. On ne peut définir la meilleure politique sans être capable de dire quels seront ses effets sous diverses hypothèses.

20. M.T. Chapalain. *L'histoire de la RCB Santé*, Bulletin RCB n° 39, Décembre 1979.

C. Rumeau-Rouquette, G. Gréban, R. Padieu, *Méthodes en épidémiologie*. Flammarion, 3^e édition 1985, p. 140 et 141.

ont été réalisées ultérieurement (1972, 1976 et 1981) dans le but de connaître l'évolution des indicateurs d'impact et d'efficacité retenus dans la politique périnatale.

2) Les liens entre contrôle de gestion et évaluation sont fréquemment mentionnés. Les deux démarches ont une parenté évidente, celle de procéder en *ex post*, le contrôle de gestion agissant toutefois plus en continu que l'évaluation. Mais elles diffèrent sensiblement par leur finalité et leurs techniques.

Sur le premier point, nous reprendrons la définition donnée par P. Gibert, P. de Lavergne²¹ : « Le contrôle de gestion a essentiellement pour but d'assurer que les résultats atteints par une organisation vont dans le sens des objectifs fixés et n'entraînent pas une consommation excessive de ressources ».

Cette définition fait clairement ressortir les différences entre contrôle de gestion et évaluation :

- l'objet analysé par le contrôle de gestion est une **organisation** ;
- l'approche est **normative** : la conduite de l'organisation donne-t-elle des résultats conformes aux objectifs ?
- elle est **interne** et concerne la gestion des ressources en fonction des objectifs fixés ou réalisés.

Une illustration de ces caractéristiques peut être trouvée dans diverses études de cas concernant les contrôleurs de gestion d'organisations publiques²².

Le contrôle de gestion apparaît ainsi comme un instrument de pilotage de la « boîte noire » constituée par l'organisation, par rétroaction des variables-objectifs observées sur les décisions. Il en résulte une prépondérance d'outils de type tableaux de bord ou indicateurs, et un faible recours à des techniques explicatives (modèles par exemple).

3) Les rapports entre évaluation des politiques publiques et contrôles administratifs ont fait l'objet de nombreuses réflexions et travaux au cours de ces dernières années²³. Cet intérêt reflète le rôle joué, en France, par ces divers contrôles et les interrogations que leur évolution suscite, compte tenu en particulier des comparaisons avec l'étranger (cf. le GAO III-2). Il traduit également l'existence de complémentarités et de proximités entre ces approches.

Avant d'en venir à ces éléments de voisinage, rappelons succinctement les points essentiels de différence entre évaluation des politiques publiques et contrôle(s) administratif(s) traditionnel(s), sans toutes les nuances qu'un exposé plus complet appellerait (voir les encadrés 3 et 4 sur la Cour des comptes et l'Inspection générale des finances²⁴.)

Différence des objets et des finalités en premier lieu. Les contrôles administratifs portent d'abord sur des **organisations** et sur la **régularité** de leur action. Si cette

21. P. Gibert, P. de Lavergne (1978). *L'analyse des coûts pour le management*, Economica.

22. E. Hachmanian, P. Hussenot (1984). *Les contrôleurs de gestion des organisations publiques*. Fondation nationale pour l'enseignement de la gestion des entreprises et Institut de management public.

23. Voir notamment B. Gournay (1983). *Tendances actuelles du contrôle des finances publiques* (Revue française d'Administration publique, n° 27) et B. Eveno (1984). *L'évaluation et le contrôle administratif* avec un commentaire par J. Chevallier (in EPP).

24. L'organisation et le rôle des diverses inspections ministérielles sont analysés de manière approfondie dans l'ouvrage de Pierre Milloz *Les Inspections générales ministérielles dans l'administration française*, 1983, Economica.

description sommaire mériterait d'être nuancée, il demeure que, par nature, les contrôles administratifs « traditionnels » regardent peu « en dehors » de l'administration, vers les résultats de l'action publique et s'intéressent prioritairement aux moyens et à leur mise en œuvre.

Différence des techniques utilisées ensuite. Dans les contrôles administratifs, l'information de base est obtenue par vérification de pièces attachées à la marche du service contrôlé et par enquêtes sur place. Le recours à des données extérieures (statistiques, sondages, etc.), avec traitement systématique, est exceptionnel. Les contraintes de délais limitent également l'utilisation croisée de plusieurs techniques d'analyse. Enfin, la relative indépendance entre opérations successives de contrôle ainsi que la rotation des responsables peuvent restreindre la comparabilité des protocoles suivis et, donc, la fongibilité des conclusions (cf. B. Eveno — op. cit.).

Malgré ces différences, il existe simultanément des éléments forts de continuité. Ils tiennent, tout d'abord, à une proximité déontologique. La séparation d'avec l'action, la postériorité de l'intervention, la possibilité de réponse contradictoire, trois principes associés à l'histoire même des corps de contrôle, ont une contrepartie directe en matière d'évaluation. On analysera ce point essentiel dans les développements consacrés à la déontologie de l'évaluation (partie III sur la mise en œuvre). Par ailleurs, on observe, sur les années récentes, un nombre croissant d'interventions dépassant le simple contrôle de régularité et rejoignant le dernier des niveaux de contrôle mentionnés dans l'article de B. Eveno²⁵ :

- niveau 1 : contrôles de régularité/honnêteté-probité ;
- niveau 2 : contrôles de régularité/respect des instructions et règles ;
- niveau 3 : contrôles d'efficacité/contrôle de gestion ;
- niveau 4 : contrôles d'efficacité collective/utilité sociale.

Reprenons, à titre d'illustration, les exemples, cités par l'auteur, d'enquêtes de l'Inspection générale des finances correspondant respectivement à ces trois derniers niveaux :

- enquête sur les services sociaux de prise en charge et placement de l'enfance malheureuse : évaluation de la politique d'aide sociale à l'enfance ;
- enquête sur la gestion des maisons de la culture : évaluation des effets redistributifs (qui bénéficie des MJC ?) ;
- enquête sur la gestion d'EDF : évaluation de la répartition de la « rente nucléaire ».

De manière générale, on constate un accroissement des missions de contrôle portant plus sur des politiques que sur des organisations (même si ce dernier aspect est présent). L'évaluation de la loi d'orientation en faveur des personnes handicapées réalisée par la Cour des comptes en est un exemple significatif²⁶, comme les autres cas mentionnés dans les encadrés 3 et 4.

4) Le tableau 1 ci-après résume de manière synthétique les caractéristiques de ces diverses approches au regard de plusieurs critères.

25. On utilise la terminologie de l'auteur. Avec un autre vocabulaire, le niveau 3 concernerait l'efficacité, le niveau 4 l'efficacité proprement dite.

26. On se reportera à la présentation faite par Pierre Brajoux dans EPP.

TABLEAU 1 : ÉVALUATION ET GENRES VOISINS

	Objet d'analyse	Situation temporelle	Finalité	Étude de l'organisation (s'il y a lieu)	Étude des effets externes	Causalité	Sources d'informations	Plan de recherche (voir chapitre IV)	Techniques utilisées	Retour sur l'action	Mise en œuvre, déontologie
Aide à la décision	Politique	<i>Ex ante</i>	Normative (optimisation)	Faible (en général)	Oui	Oui (causalités postulées)	Variables	Exceptionnellement expérimentation avec groupes témoins sinon extrapolation, simulation.	Semblables à celles de l'évaluation	Sans signification	
Contrôle de gestion	Organisation	<i>Ex tempore</i>	Normative (conformité)	Gestion des moyens	Non	Non	Internes, souvent répétitives	Pas de groupe témoin	Tableaux de bord, indicateurs	Oui en principe (pilotage)	Unité spécifique mais en général interne à l'organisation
Contrôles administratifs	Organisation et politique	<i>Ex post</i>	Normative (régularité)	Oui	Non (ou limitée)	Non	Internes, souvent spécifiques	Pas de groupe témoin (en général)	Monographie	Non intégré	Séparation contrôleur-contrôlé ; caractère contradictoire
Évaluation des politiques publiques	Politique	<i>Ex post</i>	Positive (constat)	Si nécessaire	Oui	Oui	Internes et externes, souvent spécifiques	Voir partie II	Voir partie II	Non intégré	Voir partie III (déontologie proche des contrôles administratifs)

LA COUR DES COMPTES

Juridiction administrative investie de la mission de juger les comptes, la Cour des comptes exerce à cette occasion un contrôle de la gestion des ordonnateurs sans pouvoir contraindre l'administration à suivre ses recommandations. Ainsi dotée d'attributions étendues qu'elle met en œuvre selon des modalités diverses, la Cour ne s'interdit pas, à l'occasion, l'évaluation de certaines politiques publiques.

1 — Attributions

Depuis la loi du 10 juillet 1982 qui a modifié celle du 22 juin 1967 en transférant à des chambres régionales ses attributions à l'égard des collectivités territoriales, la Cour n'est plus seul juge de droit commun des comptes publics, mais son champ de compétence reste très étendu ; il comprend :

- le contrôle juridictionnel des comptes des comptables de l'État et des établissements publics nationaux (universités par exemple) ;
- le contrôle de la gestion des organismes de sécurité sociale ;
- le contrôle des comptes et de la gestion des entreprises publiques, de leurs filiales ou sous-filiales ;
- le contrôle des organismes bénéficiaires de subventions, taxes parafiscales ou autres concours financiers de l'État ou autres collectivités publiques.

A ces quatre volets d'attributions, il y a lieu d'ajouter deux missions d'ordre général :

- la Cour est l'auxiliaire du Parlement et du Gouvernement pour le contrôle de l'exécution des lois de finances ;
- elle est tenue de procéder aux enquêtes que peuvent lui demander les commissions des finances des deux Assemblées sur la gestion des organismes soumis à son contrôle.

2 — Modalités de mise en œuvre

La compétence juridictionnelle à l'égard des comptes des comptables est d'ordre public ; la Cour doit, par ses vérifications périodiques, apurer de façon continue les comptabilités. Il en est de même à l'égard de celles des entreprises publiques dont le capital est détenu directement par l'État.

En revanche, la Cour exerce de façon facultative et *proprio motu*, selon le programme qu'elle arrête elle-même, sa compétence dans les autres domaines.

Les interventions de la Cour prennent des formes diverses. Il faut distinguer :

- les décisions en forme juridictionnelle : arrêts sur les comptes des comptables publics et comptables de fait ;
- les rapports particuliers sur les comptes et la gestion des entreprises publiques, destinés à leurs dirigeants et aux autorités de tutelle, et tenus à la disposition du Parlement ;
- les suites administratives des contrôles, généralement consacrées aux observations sur la gestion (référés du Premier président, lettres du Procureur général, propositions d'insertion au rapport public annuel).

Encadré n° 3 (suite)

3 — La Cour et l'évaluation

Défini par la loi et le règlement, le fonctionnement de la Cour est celui d'une juridiction, non d'une administration ni d'un bureau d'étude ; les chambres qui la composent sont spécialisées par catégorie d'organismes vérifiés. Toutefois, la nature des contrôles a été étendue par la loi au « bon emploi » des deniers publics et à « l'activité » des entreprises nationales, et un décret du 11 février 1985 a officialisé la possibilité de travaux relevant de la compétence de plusieurs chambres, permettant à la Cour d'aller au-delà de l'approche par organisme. Des crédits d'expertise ont par ailleurs été mis à sa disposition. L'évaluation des politiques publiques s'en trouve facilitée.

De fait, on peut citer divers exemples de politiques, analysées au niveau national ou local :

- au rapport public de 1982 :
 - « la politique en faveur des handicapés adultes » (p. 43) ;
 - « l'aide à la construction navale » (p. 110).
- au rapport public de 1983 :
 - « l'aide de l'État à l'équipement hôtelier et touristique » (p. 34) ;
 - « l'indemnisation des travailleurs sans emploi » (p. 42) ;
 - « l'exécution de la loi de programmation militaire » (p. 64).
- au rapport public de 1984 :
 - « le plan routier breton » (p. 19).
- au rapport public de 1985 :
 - « le développement agricole » (p. 15) ;
 - « la protection sociale des familles défavorisées » (p. 25).

L'INSPECTION GÉNÉRALE DES FINANCES

I — Les missions originelles

1. L'Inspection générale des finances a été créée en 1801 par un arrêté des consuls, sous le nom d'Inspection générale du trésor. Elle prend le nom d'Inspection générale des finances en 1816, lorsque le comte Corvetto décide d'y intégrer l'Inspection générale des contributions directes et du cadastre, jusque là indépendante.

En 1831, une réforme, décidée par le baron Louis, parachève l'absorption par l'Inspection générale des finances des inspections générales propres à chaque administration financière (contributions indirectes, enregistrement et domaines, douanes et postes).

2. De son origine, l'Inspection générale des finances garde trois caractères :
— elle dépend directement du ministre, au nom duquel elle exerce ses missions. Elle lui est soumise hiérarchiquement et lui adresse des rapports qui ne peuvent être communiqués à des tiers que sur sa décision ;

— elle a donc un statut de service du ministère des Finances, et non de juridiction. Elle constate mais ne prend aucune décision, ni sur le plan de la discipline ni sur le plan de la gestion ;

— elle est marquée par son statut originel de corps de contrôle ; de nombreux pans de son activité consistent toujours en vérifications de la régularité des opérations des organismes soumis à son contrôle ; au premier rang de ces organismes d'État, les services extérieurs du ministère des Finances.

3. Les pouvoirs de l'Inspection générale des finances ont été progressivement étendus :

— à des organismes exerçant des missions de service public ou dotés de régimes privilégiés (organismes de sécurité sociale, établissements de crédit à statut légal spécial, HLM, universités,...).

— aux organismes de statut privé en tant qu'ils bénéficient de concours financiers de l'État (par participation à leur capital ou par subvention) ou du produit de taxes parafiscales.

4. L'effectif budgétaire du service de l'Inspection générale des finances comprend 78 inspecteurs et 28 inspecteurs généraux ; ces derniers ont en charge la surveillance des services financiers de l'État soit au sein d'une circonscription territoriale, soit au sein d'une « division spéciale » (organismes de crédit, affaires sociales, postes et télécommunications,...).

II — Contrôle et évaluation

Le rapport qui s'instaure entre l'Inspection générale des finances et l'évaluation des politiques publiques est marqué par les caractéristiques qui viennent d'être rappelées.

1. L'évolution progressive des missions de l'Inspection

A l'extension des pouvoirs de vérification, mentionnée ci-dessus, a correspondu une extension progressive des missions de l'Inspection générale des finances, dans trois directions essentielles.

Encadré n° 4 (suite)

La vérification proprement dite s'est complétée de plus en plus d'un contrôle de gestion. Le souci qui a présidé à cette première évolution est souvent budgétaire (diminution des coûts de fonctionnement des organismes publics, ou des aides qui leur sont consenties). Mais les vérifications ont souvent pu prendre l'aspect d'audits administratifs (audit du ministère des Anciens Combattants, audit de la structure des parcs et des subdivisions du ministère de l'Équipement) voire d'analyses de systèmes complexes de gestion, notamment informatiques (analyse de projets informatiques de la Direction générale des impôts ou de la Direction de la comptabilité publique, analyse des systèmes automatiques de contrôle et de sécurité dans certains réseaux bancaires : Crédit Agricole, caisses d'épargne,...).

Le contrôle de gestion peut s'étendre à un ensemble d'organismes publics, qui sont les lieux d'application de politiques publiques : système hospitalier, structure des centres d'hébergement et de réinsertion sociale, par exemple. Les questions posées lors de semblables missions sont dès lors à la limite de l'évaluation de politiques : il s'agit de voir quelle est l'efficacité des procédures mises en place par l'administration pour garantir la gestion optimale de ces secteurs de l'activité publique, ou quelles sont les conséquences de telle ou telle modification qui a pu y intervenir (par exemple la mise en place du budget global dans le secteur hospitalier).

Les audits de procédure constituent à proprement parler des évaluations de politique publique. C'est ainsi que des études ont pu être menées sur l'efficacité des prêts spéciaux bonifiés aux investissements, sur le dispositif public de soutien aux exportations de commerce courant, sur la prime d'amélioration à l'habitat,...

2. Les atouts et les difficultés

Dans des travaux proches de l'évaluation, l'Inspection générale des finances bénéficie de deux atouts :

— l'organisation traditionnelle en brigades permet, sous l'autorité d'un responsable, d'harmoniser les investigations sur place à partir d'un échantillonnage déterminé à l'avance de manière raisonnée : cette harmonisation est évidemment nécessaire pour pouvoir tirer des enquêtes menées sur place des conclusions d'ordre général ;

— l'habitude de la vérification conduit les enquêteurs à s'assurer de la véracité et de la sincérité de l'ensemble des sources qui sont utilisées, qu'elles soient comptables ou, plus généralement, statistiques.

Il demeure qu'elle se trouvera toujours confrontée à des limites qui proviennent de la rapidité qui lui est imposée dans l'exécution de ses tâches, du faible nombre de ses effectifs et de leur forte rotation, et plus généralement de l'optique financière qu'elle adopte généralement. Cette dernière caractéristique, si elle peut apparaître comme génératrice d'une certaine « étroitesse » pour ceux qui ont de l'évaluation une conception plus extensive et pluridisciplinaire, demeure néanmoins une garantie de sérieux et de professionnalisme ; elle limite les risques d'évaluation-justification et n'empêche pas, à l'expérience, de porter, sur les domaines qui le méritent, une appréciation plus large tenant à la réalisation des objectifs visés et à l'adéquation des moyens mis en place pour y parvenir.

I — 4 Évaluation des politiques et analyses de l'activité publique

L'évaluation des politiques publiques, en tant que pratique, renvoie donc à plusieurs genres voisins. A un plan plus conceptuel, elle peut être également située par rapport aux **principales problématiques de l'activité publique**²⁷. Une analyse de « positionnement », même schématique, n'est pas inutile, compte tenu des débats qui opposent souvent les tenants de ces diverses écoles et qui obscurcissent parfois la spécificité de l'évaluation proprement dite.

1) **La première problématique**, la plus traditionnelle, consiste à considérer l'État comme une institution, c'est-à-dire une structure sociale spécifique dotée de prérogatives et d'un mode de fonctionnement différents du reste du corps social, lui-même caractérisé par l'opposition des intérêts privés. Dans cette perspective, l'administration est distincte du pouvoir politique dont elle est l'instrument. Elle est distincte également de la société civile à laquelle elle assure maintien de l'ordre, application des lois et fonctionnement des services publics.

Cette approche, qui a beaucoup régressé dans les pays anglo-saxons, est restée forte en France, ce d'autant plus que dans le conflit idéologique qui écartèle les sciences sociales, l'analyse marxiste partage parfois avec l'analyse libérale une vision instrumentale de l'administration.

La seconde problématique est de nature économique : l'activité publique est vue comme une activité de **production et de distribution de biens et de services collectifs**. Elle procède donc à la recherche des niveaux optimaux de productions, à la mesure de l'adéquation de cette production à la demande, etc.

Une troisième problématique consiste à voir les administrations publiques comme des **organisations** de caractéristiques générales semblables à celles des entreprises ou autres groupes sociaux finalisés. La **sociologie des organisations**, la **recherche en gestion** reconnaissent dans les administrations l'existence de structures ayant une autonomie relative, fonctionnant selon un régime de rationalité limitée et où les objectifs officiels doivent compter avec les stratégies des acteurs. Sur le plan des méthodes, cette démarche se caractérise par une **attitude inductive** fondée sur l'observation et sur l'enquête de terrain.

Enfin, l'analyse des politiques publiques proprement dite appréhende l'État en action, à travers les processus par lesquels les instances politico-administratives agissent dans et sur la société.

C'est avec cette dernière problématique que l'évaluation des politiques publiques entretient les liens les plus directs. A certains égards, l'évaluation est une composante de l'analyse des politiques publiques, qui fait elle-même de multiples emprunts aux trois premières approches.

27. Voir J.P. Nioche, *Science administrative, management public, analyse des politiques publiques*, Revue française d'Administration publique, n° 24, octobre-décembre 1982 et X. Greffe, *L'évaluation des activités publiques*, Recherches économiques et sociales, 1^{er} et 2^e trimestres 1985.

2) **En tant que participant de l'analyse des politiques publiques, l'évaluation est à la fois une étape du processus d'action des Pouvoirs publics et une des tâches de l'analyste qui cherche à comprendre tout ou partie de la séquence suivante :**

— la formation de la politique : comment et à qui la situation apparaît-elle problématique ? Quelles réponses possibles sont élaborées ? Par qui ? Pourquoi ? Quels intérêts sont en jeu ? Comment une décision (ou non-décision) est prise finalement ? Pourquoi cette décision-là et à ce moment-là ?

— la mise en œuvre de la politique : quels sont les moyens, les procédures ? Quelles organisations, quels intervenants régissent le processus ? Comment les choses se passent-elles ? Pourquoi ?

— l'évaluation de la politique : quels effets prévus a-t-elle produits sur la situation visée, sur d'autres domaines ? Quelles modifications se sont produites dans l'équilibre des pouvoirs ? etc.

Une première caractéristique de l'analyse des politiques publiques est de faire disparaître les barrières entre système politique, système administratif et ensemble de la société et d'opérer une coupe transversale. Toute politique comporte, dans des proportions variables, une dimension managériale (résoudre un problème, réaliser un projet) et une dimension politique (maintenir ou modifier une certaine situation de pouvoir). L'analyse se développera donc plus ou moins sur ces deux plans interdépendants, étant entendu que la dimension politique peut concerner à la fois le système de pouvoir spécifique dans lequel se forme et se déploie la politique, et le système politique global dans lequel celui-ci s'insère. Cette **multidimensionnalité** d'étude se retrouve en matière d'évaluation.

Deuxième caractéristique : l'analyse de politiques publiques fait souvent appel à une démarche diachronique. Celle-ci ne doit pas être confondue avec une simple chronologie des faits, l'ordre des étapes présentées ci-dessus pouvant être bouleversé par les stratégies des acteurs : anticipations, politique appliquée sans que la décision formelle soit prise, ou au contraire, jamais appliquée, etc. On verra là aussi que la **dimension temporelle** est centrale en matière d'évaluation.

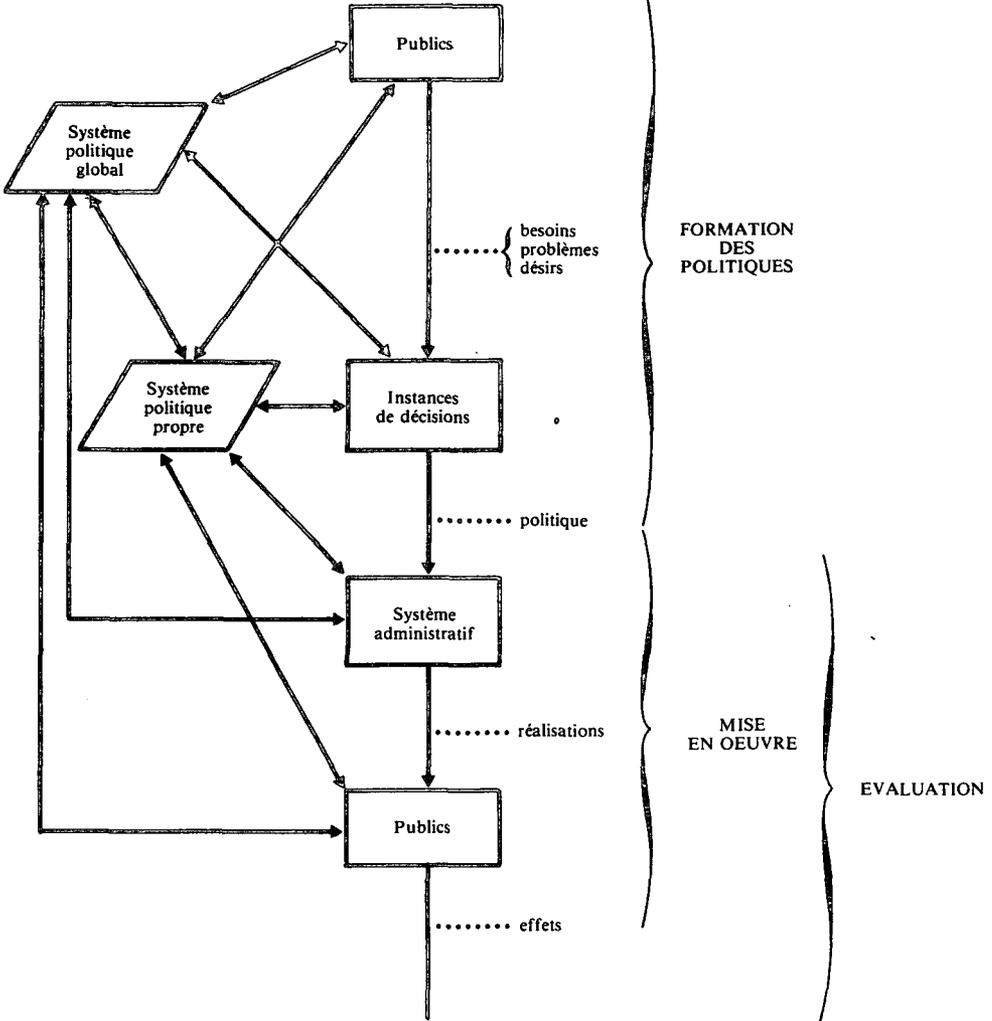
Enfin, l'analyse des politiques publiques a généralement une ambition explicative et pas seulement descriptive. Elle suppose une conceptualisation des processus en cause et la définition d'une méthodologie permettant de générer et de tester des hypothèses causales. Les méthodologies sont variables et recourent à l'arsenal des diverses sciences sociales, économie, science politique, sociologie... Au-delà de ces emprunts, l'analyse de politiques a donné lieu à la mise au point d'un important corpus de méthodes propres, en matière de plans de recherche comme de techniques de collecte et de traitement des données (cf. II infra). Cette **problématique explicative** et ces méthodes sont partie intégrante de la démarche évaluative.

L'encadré n° 5 permet de visualiser (de façon simplificatrice) les principaux éléments de l'analyse de politiques publiques et d'y situer l'évaluation.

3) Si l'évaluation est, pour partie, une composante de l'analyse des politiques publiques, qui comporte l'étude empirique de leurs effets, elle déborde cet objet et joue un rôle spécifique au plan conceptuel pour deux raisons.

Encadré n° 5

Les éléments de l'analyse des politiques publiques



La première tient au fait précédemment mentionné que, souvent, l'évaluation impose de ne pas se limiter à la mesure des effets, mais de pénétrer dans la « boîte noire », d'analyser les organisations et les procédures de mise en œuvre de la politique, ou même de reconstituer la genèse de la politique. Ces cas ne sont pas rares. Par exemple, en matière de politiques sociales, il peut être aussi difficile de comprendre les actions réelles accomplies par de multiples intervenants que de mesurer leurs résultats.

De telles situations sont parfois mises en exergue pour soutenir que les politiques publiques ne seraient pas évaluables. Selon cette thèse, l'évaluation, pour être pertinente, exigerait que la politique ait été décidée rationnellement et avec une définition explicite d'objectifs. En fait, la mesure empirique des effets d'une politique peut être conduite, que cette politique soit le produit d'un choix rationnel, de mécanismes bureaucratiques ou d'un compromis politique. La décision incrémentale²⁸, c'est-à-dire l'élaboration d'une politique au cours du temps par tâtonnement progressif, conduisant à des actions sédimentaires, peut rendre difficile la mesure des effets. Mais le modèle incrémental n'est qu'une situation particulière de décision politique²⁹. Ces cas limites viennent rappeler que toutes les politiques ne sont pas « également » évaluables. L'analyse de la formation et de la mise en œuvre de la politique peuvent alors servir de substitut à une évaluation³⁰, ou, à tout le moins, en constituer un préalable indispensable.

La deuxième raison du rôle central de l'évaluation dans l'étude des politiques publiques est qu'elle recoupe largement les problématiques décrites ci-dessus. Chacune d'entre elles fournit un cadre conceptuel et des méthodes à l'évaluation qui, en tant que **discipline appliquée, vit d'emprunts**. Dans chaque approche l'évaluation est définie par une question spécifique³¹, sous forme normative ou positive :

— pour la problématique institutionnelle (« la loi est-elle respectée ? ») : répondent à cette question le magistrat au plan normatif, et, de façon positive, le politologue qui enquête sur la non parution des décrets d'applications d'une loi votée ;

— pour la problématique des biens (« les administrations fournissent-elles correctement biens et services collectifs ? ») : répondent à cette question la démarche normative de l'analyse coûts-avantages ou la démarche positive de l'enquête de consommation ou de la mesure d'indicateurs sociaux ;

— pour la problématique organisationnelle (« les administrations chargées de la politique agissent-elles avec cohérence, efficacité et efficacie ? ») : l'analyse économique de la bureaucratie répond de façon normative à cette question, la sociologie des organisations de façon positive ;

— pour la problématique des politiques publiques (« quels sont les effets attribuables à la politique ? ») : la « policy science » répond de façon normative à

28. C. Lindblom. *The science of muddling through*.

29. J.P. Anastassopoulos, G. Blanc, J.P. Nioche, B. Ramanantsoa, *Pour une nouvelle politique d'entreprise*, PUF, 1985, 3^e partie.

30. F.G. Padioleau, *Quand la France s'enferme*, PUF, 1981.

31. X. Greffe, op. cité, p. 11.

cette question, l'évaluation de façon positive, en utilisant les répertoires de méthode de différentes approches et des sciences fondamentales qui les sous-tendent.

4) Sur ce dernier point, les progrès de l'évaluation, aux États-Unis notamment, ont été en effet soutenus par la reconnaissance croissante des normes de scientificité des sciences « dures » dans les sciences « molles ». L'utilisation, en psychologie par exemple, de l'analyse des données, de la modélisation, de l'expérimentation, a joué un grand rôle dans le progrès en crédibilité des sciences humaines auprès des autres scientifiques comme auprès des décideurs. Le débat sur la nature plus ou moins scientifique de la démarche évaluative, sur la capacité des évaluateurs à échapper aux biais idéologiques et aux choix de valeurs, est vif entre chercheurs. Une voie consiste à reconnaître le caractère contradictoire par nature des points de vue qui peuvent s'exprimer sur les résultats d'une politique publique et à adopter une démarche organisant la confrontation des avis. Une telle démarche peut être constituée de plusieurs évaluations menées en parallèle et donnant lieu à une analyse combinée. A la limite, la procédure de confrontation des points de vue devient la seule méthode et il est espéré qu'un consensus s'établira qui sera l'évaluation. Des politiques conflictuelles, à forte connotation idéologique, pour lesquelles il n'existe pas d'instruments de mesures acceptés par tous, peuvent justifier le recours à de telles procédures. Il s'agit là de domaines dans lesquels l'évaluation ne peut être qu'interprétée *largo sensu*, mais constituer le domaine de prédilection de l'analyse clinique et de l'étude de cas.

I - 5 L'évaluation des politiques publiques : une approche spécifique

Au total, les développements précédents ont permis de reconnaître les ressemblances et les différences entre l'évaluation des politiques publiques et des pratiques voisines, comme de marquer sa position par rapport aux diverses problématiques de l'activité publique.

Il en ressort que la définition stricte de l'évaluation des politiques publiques proposée ne correspond pas à une problématique d'exclusive. On voit ainsi que, d'une part, l'évaluation des politiques publiques ne saurait se substituer, pour certaines questions et certains objectifs, à d'autres approches qui gardent entièrement leurs droits, que, d'autre part, dans une situation concrète et en fonction des conditions d'environnement, l'usage conjoint de diverses problématiques peut être indispensable. Ainsi, l'évaluation des politiques publiques ne peut rendre les mêmes services de pilotage que le contrôle de gestion. Dans le cas d'organisations publiques, il serait évidemment aberrant de prétendre substituer aux vérifications de régularité la seule démarche évaluative. En outre, un tel contrôle peut être un préalable nécessaire à la réalisation d'une évaluation

pertinente ou, à tout le moins, constituer un appui très utile³². De même, en ce qui concerne le cycle politique, qu'on passe de l'*ex post* proprement dit à des problèmes de formulation de politique, alors les démarches associatives retrouvent leurs droits.

Le développement de l'évaluation des politiques publiques requiert de rejeter toute prétention à une universalité factice. C'est en la fondant sur une conception précise, en l'articulant autour de quelques principes clairs, méthodologiques et déontologiques, qu'on peut en espérer une contribution spécifique, qui complète efficacement d'autres apports.

32. On notera ainsi dans la partie IV que la situation de départ pour le développement de l'évaluation est plus confortable dans les domaines où « l'interne » pèse d'un poids déterminant. Dans ce cas en effet l'apport des contrôles administratifs peut être proche, pour certaines questions, de ce que donnerait une véritable évaluation (exemple : entreprises publiques). *A contrario*, vérifier la régularité d'application d'une réglementation ne fournit que peu de renseignements sur son efficacité externe.

II

Les méthodes de l'évaluation



II

Les méthodes de l'évaluation

II - 1 Quelques éléments de terminologie

Nous abordons dans cette partie le cœur « technologique » du rapport à travers la présentation des plans de recherche et des techniques de l'évaluation. Il est utile, en préambule, de rappeler l'objectif visé et d'introduire un minimum de terminologie¹.

Comme nous l'avons dit en introduction, les travaux du groupe n'avaient pas d'ambition novatrice fondamentale au plan technique, pour deux raisons. D'une part, l'acquis est de bonne qualité. Il a donné lieu à plusieurs présentations académiques, notamment dans la littérature anglo-saxonne². D'autre part, grâce à ce capital, il **n'existe pas de goulot d'étranglement technique** pour le développement de l'évaluation en France, au moins en ce qui concerne la boîte à outils. L'affinement des méthodes existantes n'est donc pas une priorité. Par contre, ce rapport garde une fonction importante de diffusion et de discussion critique de ces méthodes. En outre, il consacre à la modélisation des développements spécifiques, peu fréquents dans la littérature usuelle.

La filiation expérimentale et psychologique de celle-ci s'exerce en défaveur de l'approche modélisatrice, plus développée en économie ou dans les sciences physiques. Cette approche présente un intérêt particulier pour de nombreux problèmes d'évaluation à fort contenu économique.

La présentation distingue entre **plans de recherche et techniques d'exploitation de l'information**³.

Un **plan de recherche** (protocole) spécifie une **organisation méthodologique générale** de l'évaluation en fonction de **l'information** empirique (observations) disponible ou à collecter.

1. Un glossaire général figure en fin de rapport.

2. Les synthèses les plus commodes sont fournies par les articles de D.T. Campbell (*Assessing...*) et T. J. Cook, D. T. Campbell (*The design...*). Voir également la présentation de R. Angelmar (EPP) ainsi que :

— E. P. Scioli, T. J. Cook *Methodologies for analyzing public policies* (1975, Lexington books).

— S. S. Nagel *The policy studies handbook* (1980, Lexington books).

3. Cette distinction entre plans de recherche et techniques, retenue pour la simplicité de l'exposé, est, à certains égards, trop tranchée : ainsi, l'étude de cas ou la modélisation, que nous classerons comme techniques, peuvent-elles s'apparenter à des plans de recherche spécifiques.

Une technique d'exploitation de l'information consiste en un outil particulier de dépouillement, de traitement, de synthèse et d'interprétation. Une même technique (par exemple, une procédure particulière d'analyse de séries temporelles) pourra évidemment être utilisée dans plusieurs plans de recherche.

Le classement des plans de recherche retenus ci-après prend en compte leur caractère « expérimental » plus ou moins « pur », c'est-à-dire la présence, ou non, d'un groupe témoin équivalent (ou à peu près équivalent⁴). On distinguera ainsi :

— l'évaluation sans groupe témoin (II-2) : les plans de recherche ne font pas appel à la comparaison de groupes affectés et de groupes non affectés par la politique ;

— l'évaluation par comparaison avec un groupe témoin non équivalent (II-3) : les plans de recherche s'appuient sur une comparaison avec des groupes non affectés (ou affectés différemment) par la politique. Toutefois, on ne peut être assuré que les caractéristiques de ces groupes sont totalement identiques à ceux du groupe affecté par la politique ;

— l'évaluation par comparaison avec un groupe témoin équivalent ou « expérimentation vraie » (II-4) : ce cas correspond au schéma « idéal » du paragraphe 1-2. L'équivalence du groupe non affecté doit être comprise dans un sens statistique : le groupe traité et le groupe témoin réputé équivalent doivent être représentatifs de la population totale, et sont en général constitués par tirage au sort.

Un sort particulier est fait à l'approche par modélisation. La modélisation, conçue de manière générale, occupe une position intermédiaire entre plan de recherche et technique. En réservant le terme à la construction d'une représentation formalisée, susceptible d'estimation numérique, à partir d'une théorie *a priori* explicative⁵, on se rapproche d'un plan de recherche spécifique. On discutera de la position de la modélisation par rapport aux problèmes d'évaluation dans le paragraphe II-5 « évaluation et théorie : de l'étude de cas à la modélisation explicative ».

Enfin, cet exposé analytique se conclut par une mise en perspective synthétique des divers plans de recherche (II-6).

Une dernière remarque. La présentation des divers plans de recherche a recours aux notations introduites par T. J. Cook et D. T. Campbell (*The design... - op. cit.*) :

- X désigne la mise en œuvre de la politique ;
- O désigne une observation réalisée sur un groupe ;
- les indices 1, 2, ..., n se réfèrent à l'ordre séquentiel de mise en œuvre des politiques (X_1, \dots, X_n) et des observations correspondantes (O_1, \dots, O_n).

4. Ce terme est précisé ci-après. Voir également le glossaire.

5. La « théorie » n'est évidemment jamais totalement absente des démarches ne recourant pas à la modélisation. On dit que l'évaluation requiert en général une reconstruction du réel, avec des opérations de coupure qui supposent donc une conceptualisation minimale *a priori* en matière de liaisons causales (on exclut *a priori* un « modèle » du style « tout est dans tout »).

II - 2 L'évaluation sans groupe témoin

Dans l'évaluation sans groupe témoin l'évaluateur n'a d'information (observations) que sur le groupe affecté par la politique évaluée. Il doit se contenter d'examiner la situation après mise en œuvre de la politique, en la comparant, au mieux, avec celle qui prévalait auparavant. La validité tant interne qu'externe de ces plans de recherche est donc, en règle générale, peu satisfaisante en raison de la difficulté d'isoler la politique des autres sources possibles d'évolution. Cette situation est toutefois caractéristique d'un grand nombre de cas d'évaluation, soit pour des mesures de politique ayant une portée générale, soit, au contraire, pour des mesures extrêmement spécifiques rendant délicates des comparaisons externes. On distinguera ci-après trois plans de recherche :

- l'observation après mise en œuvre de la politique ;
- l'observation du même groupe avant et après ;
- le suivi de la mise en œuvre progressive de la politique, c'est-à-dire une variante, en continu, du plan de recherche précédent.

1) L'observation du seul groupe concerné, après la mise en œuvre de la politique, constitue le plan de recherche le plus rudimentaire : il exclut toute comparaison avec la situation qui prévalait avant application de la politique.

Avec les notations précédentes, il est représenté par le schéma 1.

Groupe cible : X 0

Schéma 1 - Observation après mise en œuvre de la politique

Ce plan de recherche n'est approprié que si les objectifs poursuivis au travers de l'évaluation sont limités. Il peut permettre :

— de vérifier que la politique a bien été mise en œuvre et de constater de quelle manière elle l'a été. Ce type d'évaluation est effectué de façon assez systématique pour les moyens, plus fréquemment pour les réalisations mais rarement pour l'impact de la politique dans le public. Ainsi l'évaluation des préretraites a notamment porté sur le respect de l'engagement de maintien des effectifs, souvent par les entreprises choisissant d'adhérer au dispositif⁶. De même, une analyse de conformité aux règles d'attribution a été incluse dans l'étude prioritaire RCB sur l'efficacité des prêts bonifiés à l'agriculture⁷ ;

— de mesurer les effets « comptables » immédiats de la politique (nombre de personnes concernées, coût financier, etc.) sans chercher à fournir un schéma explicatif des phénomènes observés. On peut citer par exemple le suivi statistique de la mise en place des travaux d'utilité collective (TUC) (encadré n° 1) et des mesures de préretraites⁶ ;

6. Voir la discussion et les références contenues dans l'étude de M. Elbaum et A. Fonteneau (1984) *Les politiques de l'emploi face à la crise* (ronéo OFCE).

7. Voir D. Gagey et P. Lenoël (1981) *Bilan de l'étude prioritaire sur l'efficacité des prêts bonifiés à l'agriculture* (Revue RCB, n° 44).

— d'améliorer la compréhension des comportements des groupes concernés par la politique, voire d'en dégager une typologie. Ainsi, dans le cadre de l'étude sur les aides à l'élevage bovin, une enquête a été réalisée dans trois départements, auprès d'acteurs de la vie locale, afin de mettre en évidence les facteurs d'inefficacité de l'action publique liés à une mauvaise perception des attentes et des comportements de ces acteurs⁸. Des typologies ont pu être également établies dans différentes études d'évaluation : réaction des entreprises face à la réduction du temps de travail⁹ ou aux mesures de préretraites, motivation des organismes employant des TUC (encadré n° 1) par exemple. Dans un domaine différent, le bilan du TGV Sud-Est réalisé par la SNCF a permis d'ajuster les modèles de prévision de trafic et donc d'améliorer les bilans prévisionnels concernant les TGV en projet.

8. Voir la communication présentée au groupe par C. Le Pen, *Présentation méthodologique de l'étude d'évaluation de la politique publique de l'élevage bovin* (1985, université Paris IX Dauphine, ronéo).

9. Voir M. Pepin, D. Tonneau *Réglementation sociale et vie des entreprises : mise en œuvre des ordonnances de janvier 1982 sur les 39 heures et la 5^e semaine de congé* (1984, communication aux 4^{es} Journées d'économie sociale, ronéo). Cette étude est résumée dans l'encadré n° 8.

L'ÉVALUATION DE LA MISE EN PLACE DES TRAVAUX D'UTILITÉ COLLECTIVE

Les travaux d'utilité collective (TUC) sont des stages de 3 à 12 mois dans des organismes du secteur non marchand s'adressant à des jeunes disposant en général d'un faible niveau de formation. Ils sont rémunérés à raison de 1 200 F/mois à la charge de l'État auxquels peut s'ajouter une indemnité de 500 F/mois à la charge de l'organisme d'accueil. Les objectifs des pouvoirs publics étaient au départ de mettre en place 100 000 TUC, chiffre porté à 200 000 en janvier 1985.

Le suivi de la mise en place des TUC consiste en une **évaluation de réalisation ex post** sans groupe de contrôle. Il a nécessité un système d'information statistique, *ad hoc*, le plus lourd mis en place pour une série de mesures de politique d'emploi :

— l'information transmise par téléphone par les directions départementales du travail et de l'emploi est traitée par informatique de manière à fournir une **statistique hebdomadaire provisoire** (mission TUC) ;

— une **statistique mensuelle** est en outre traitée par le Service des études et de la statistique du ministère du Travail et de l'Emploi, à partir des bordereaux nominatifs concernant respectivement les jeunes et les organismes.

En outre certaines opérations d'évaluation ont d'ores et déjà été entreprises :

— une investigation a été menée dès le début de la mise en place du dispositif par le Service des études et de la statistique du ministère du Travail et de l'Emploi auprès d'environ 60 organismes d'accueil¹. Elle a donné lieu à des monographies, qui montrent d'une part que les tâches faisant l'objet de TUC ne pouvaient être remplies auparavant, d'autre part que l'aspect qualification est réduit ;

— une monographie concernant les résultats dans la région Rhône-Alpes est menée à l'initiative du SES ;

— une étude de la Direction de la prévision a cherché à déterminer les conséquences des TUC sur les autres types de stages dont peuvent bénéficier les jeunes concernés ;

— une enquête par voie postale a été réalisée par le Centre d'études sur l'emploi auprès de 2 000 organismes d'accueil. Une précédente enquête portant sur 30 organismes seulement avait permis de déterminer 4 attitudes de leur part vis-à-vis des TUC : le bénévolat rémunéré, l'occupation de surnuméraires, le salariat précaire, le stage en formation alternée ;

— un sondage a été réalisé par l'IFOP afin de connaître la situation, après trois mois, de jeunes sortis du dispositif à la mi-juillet 1985.

Enfin, le SES et la Délégation à l'emploi, avec la collaboration technique au Centre d'Études et de recherches sur les qualifications, lanceront une enquête sur le devenir des jeunes sortis des TUC.

1. Voir : Travail et Emploi - Décembre 1985.

Si le plan de recherche « observation après mise en œuvre de la politique » peut suggérer des liens de causalité, la validité interne ou externe des conclusions tirées demeure fragile et les possibilités d'extension à d'autres situations très limitées. D'une part, l'absence de comparaison avec la situation antérieure à la mise en œuvre de la politique ne permet pas de vérifier si celle-ci a véritablement coïncidé avec une inflexion de l'évolution de certains indicateurs. D'autre part, l'absence du groupe témoin empêche d'identifier avec certitude ce qui se serait passé « si la politique n'avait pas été appliquée ».

Les techniques utilisées sont, d'une part l'étude de cas, d'autre part des techniques exploitant des informations issues d'enquêtes statistiques (analyse des données, économétrie sur coupes transversales).

L'étude de cas est de pratique fréquente pour l'évaluation *ex post* sans groupe de contrôle. Selon la terminologie du « General Accounting Office », elle consiste en « une description analytique d'un événement, d'un processus, d'un organisme ou d'un programme »¹⁰. Elle est souvent retenue lorsque les mécanismes mis en œuvre par la politique étudiée sont complexes et étalés dans le temps, avec d'éventuels phénomènes de sédimentation¹¹. L'étude de cas peut avoir un caractère seulement descriptif (quelles ont été les modalités de mise en œuvre de la politique ?) ou au contraire permettre d'avancer des explications quant aux effets de la politique (sans pour autant les valider). La possibilité de dégager un schéma explicatif peut se trouver accrue lorsque l'on multiplie les études de cas sur une politique ou sur des politiques voisines¹². Toutefois, le fait que les cas étudiés ne résultent pas d'un échantillonnage statistique limite la validité d'une telle approche. En particulier il serait imprudent de généraliser à l'ensemble de la population des constats réalisés sur des unités économiques de caractéristiques très spécifiques. En outre, certaines difficultés inhérentes à l'étude de cas doivent être prises en compte dès sa phase de conception : risques de subjectivité liés à l'utilisation d'informations fortement qualitatives, fiabilité médiocre des données quantitatives si la collecte n'est pas suffisamment structurée et systématique. Il reste que l'étude de cas est souvent un préalable indispensable à l'application de plans de recherche plus raffinés ou de techniques plus dures, sauf à courir de grands risques de naïveté.

La deuxième source d'informations utilisable pour l'évaluation *ex-post* sans groupe de contrôle est l'enquête statistique. Par rapport à l'étude de cas, elle présente l'avantage de reposer sur l'analyse d'échantillons de la population concernée, plus représentatifs et généralement plus nombreux. Elle suppose qu'un échantillonnage a pu être effectué préalablement à l'évaluation, ce qui n'est possible que si l'on dispose d'une bonne connaissance statistique des groupes visés par la politique. Dans l'étude sur l'évaluation des aides à l'emploi, une enquête statistique a ainsi été réalisée en fonction de la taille des entreprises¹³. L'application de techniques

10. Voir *Designing evaluations* (1984, GAO, Methodology transfer paper 4).

11. Voir par exemple les études de cas de politiques industrielles mentionnées en II-2. Le cas limite dans ce domaine est celui où l'évaluateur se mue en historien et recherche quelle a été véritablement la politique suivie.

12. Par exemple la multiplication d'études de cas de politiques industrielles peut permettre de suggérer quelques grandes lois ou mécanismes. Voir M. Bauer, E. Cohen. *Les grandes manœuvres industrielles* (1985, Calmann-Lévy).

13. *Image et impact des aides à l'emploi dans les entreprises* (juin 1985, Louis Harris-France, ronéo).

d'analyse des données aux résultats de l'enquête statistique peut mettre en évidence une typologie de comportements vis-à-vis de la politique, de manière beaucoup plus fiable que l'étude de cas qui permet seulement de suggérer des comportements différenciés. L'enquête sur les aides à l'emploi a établi que les grandes entreprises ont tendance à négocier les aides qu'elles utilisent (en recourant à un mélange d'aides standard) alors que les entreprises plus petites utilisent des aides standard au coup par coup¹³.

L'enquête statistique peut enfin permettre, sous certaines conditions, de tester des liens de causalité. Ceci suppose une modélisation microéconomique préalable et un échantillon suffisamment hétérogène pour que l'estimation économétrique sur coupes instantanées puisse isoler les effets de la politique en tirant parti des différences entre les observations. Lorsqu'il existe des contraintes de délais serrées, cette approche présente toutefois le défaut d'être lourde et délicate du point de vue des techniques utilisées.

2) L'observation du même groupe avant et après constitue le deuxième plan de recherche utilisable en l'absence de groupe de contrôle. Dans ce cas, le groupe cible, observé avant mise en œuvre de la politique, fait l'objet d'une observation après mise en œuvre.

Ce plan de recherche, appelé fréquemment étude longitudinale, est représenté par le schéma n° 2.

Groupe cible : $O_1 \times O_2$

Schéma n° 2 — L'observation avant-après

Il est répandu, ne serait-ce que parce qu'il correspond à une démarche naturelle en matière d'évaluation : le premier réflexe de l'évaluateur est de vérifier que la période d'application de la politique a coïncidé avec une modification de certains indicateurs (même si la constatation de cette concomitance ne permet pas de déduire l'existence d'un lien de causalité)¹⁴. Ce plan de recherche permet en outre des présentations graphiques très parlantes (rupture de courbe). De tels plans de recherche avant-après ont été ainsi appliqués en France pour le cas des politiques de sécurité routière (voir l'exemple de l'encadré n° 2).

Il convient toutefois de bien reconnaître les limites de validité de ce plan de recherche commode.

Elles résident principalement dans la difficulté de discerner les changements dus à la politique et ceux résultant de la modification de paramètres extérieurs. Cette difficulté sera d'autant plus grande que l'on retient des indicateurs globaux ne permettant pas de tirer parti de la variance pouvant exister au sein de la population observée (le groupe observé n'est pas, en effet, nécessairement homogène).

14. L'article de D. T. Campbell (*Assessing et...* - op. cit.) donne de nombreux exemples (avec graphique) de telles études avant-après : impact de la « législation pornographique » au Danemark sur l'évolution des délits sexuels, loi sur l'interruption de la grossesse en Roumanie et naissances, contrôles anti-alcooliques en Grande-Bretagne et accidents routiers, etc.

Encadré n° 2

**UN EXEMPLE DE PLAN DE RECHERCHE AVANT/APRÈS :
L'ÉVALUATION DES ACTIONS DE SÉCURITÉ ROUTIÈRE ET
D'EXPLOITATION DE LA ROUTE**

La méthode utilisée consiste à calculer le taux (apparent) d'efficacité d'une action. Celle-ci (e) se définit de la façon suivante :

$$e = 1 - \frac{N_2}{N_1(1+g)}$$

où N_1 = nombre d'accidents/an avant aménagement

N_2 = nombre d'accidents/an après aménagement

g = taux de croissance du trafic pendant la période considérée.

Les évaluations réalisées au cours des dernières années ont donné les résultats suivants :

		Taux d'efficacité
Zones d'accumulation d'accidents (étude réalisée en 1976)	carrefours	66 %
	virages	77 %
	écrêtements	72 %
	glissières	50 %
	abattages d'arbres	56 %
	passages dénivelés	90 %
	ensemble	70 %
Carrefours giratoires		77 % (accidents) 94 % (accidents mortels) 82 % (blessés)
Carrefours à feux (rase campagne)		29 % (accidents) 60 % (accidents mortels) 26 % (blessés)

Un calcul de rentabilité est également réalisé compte tenu du coût de l'opération, du taux apparent d'efficacité et d'une valeur attribuée à la vie humaine.

Les comparaisons avant-après menées à l'occasion du bilan du plan textile (voir encadré n° 11) mettent en évidence ces difficultés pour l'établissement de liens de causalité : comment distinguer effets à court terme et tendance, comment éliminer l'influence de l'environnement international ou de la politique économique générale pour ne retenir que celle attribuable au plan ?

De même, les comparaisons des mesures préventives avant et après 1973 pour les accidents de la circulation en France sont (partiellement) rendues délicates par la diminution relative du trafic routier liée à la crise de l'énergie.

L'existence d'effets d'une politique très différents à court terme et à moyen terme peut également limiter la portée d'une comparaison avant-après. Par exemple, l'évaluation du plan textile a été réalisée pendant la phase de mise en œuvre du plan : il ne pouvait donc être question d'analyser ses effets à long terme sur l'investissement et la situation financière des entreprises. Dans ce cas, il peut être nécessaire de réaliser des observations à différentes échéances pour avoir une appréciation complète des effets de la politique¹⁵. On reviendra sur ce point dans la discussion du plan de recherche du type observatoire permanent.

En outre, un plan de recherche avant-après doit être utilisé avec précaution dans le cas de politiques visant à prévenir des **phénomènes aléatoires**. C'est par exemple le cas de politiques d'infrastructure en matière de sécurité routière. Leurs effets ne peuvent être correctement appréciés que si l'on dispose d'un recul suffisant. Ce sera *a fortiori* le cas pour les politiques de protection civile destinées à prévenir des catastrophes naturelles dont le caractère aléatoire n'apparaît que sur longue période (crues trentenaires, etc.).

Comment améliorer la validité de plans de recherche avant-après ? Tout d'abord les **recoupements** que permet l'utilisation de plusieurs techniques d'évaluation fournissent un moyen d'éprouver la robustesse des résultats. Si les différentes techniques utilisées conduisent à des résultats concordants, la validité de l'évaluation s'en trouve immédiatement renforcée. A l'opposé, des conclusions discordantes incitent à la prudence vis-à-vis des enseignements de l'évaluation. Ainsi, la mesure de la compensation salariale consécutive à la réduction du temps de travail à 39 heures a été réalisée de trois façons différentes : à partir d'une enquête auprès d'un échantillon d'entreprises, à partir d'une analyse tendancielle d'évolution des salaires et à partir de l'estimation d'une relation économétrique. Ces trois évaluations ont donné des réponses divergentes puisqu'elles aboutissent à des compensations salariales allant de 0 à 70 %¹⁶.

Une autre manière d'améliorer la validité d'une évaluation avant-après consiste à examiner l'effet de la politique sur des unités microéconomique suffisamment différenciées (**coupe transversale**). L'idéal est alors de construire un modèle microéconomique et de le tester économétriquement. On reviendra sur cette démarche de modélisation microéconomique pour le cas de l'allocation de parent isolé, où elle a été utilisée dans un plan de recherche avant-après (encadré n° 10).

15. Cette remarque s'applique à de nombreux domaines : politique industrielle (les effets conjoncturels l'emportent à court terme sur les réformes structurelles), éducation, santé...

16. Voir M. Elbaum, A. Fonteneau (op. cit.).

3) Par extension de l'observation avant-après, on aboutit au troisième plan de recherche utilisable en l'absence de groupe de contrôle : le suivi de la mise en œuvre progressive de la politique et de la production progressive de ses résultats. Ce type de plan de recherche est représenté sur le schéma n° 3.

Groupe cible : O₁ X₁ O₂ X₂ O₃ X₃...

Schéma n° 3 - Le suivi de la mise en œuvre progressive de la politique

Ce schéma peut correspondre :

— soit au cas d'un observatoire permanent, comme ceux mis en place dans le domaine de la formation ;

— soit au cas d'une politique, constante dans ses objectifs, mais progressivement réajustée dans ses moyens, au vu des résultats successifs d'évaluation.

L'observatoire permanent consiste à observer les effets de la politique soit à différents stades de sa mise en œuvre, soit au cours du temps.

Le dispositif mis en place par le ministère de l'Éducation nationale¹⁷ (Service de la prévision, des statistiques et de l'évaluation), pour évaluer les programmes scolaires des différentes classes, permet ainsi, en travaillant sur les mêmes échantillons d'établissements et en décalant les évaluations dans le temps (1979 pour les programmes du CP, 1981 pour ceux du CE2, etc.), de suivre certains élèves dans leur scolarité (encadré n° 3). Ce dispositif a notamment été utile :

— pour informer les responsables du ministère du niveau atteint par les élèves. Ces éléments ont pu être utilisés pour l'élaboration des nouveaux programmes et surtout pour la rédaction des annexes et compléments à ces programmes. Ce retour sur les objectifs (soit pour les infléchir, soit pour les confirmer) est en effet un des résultats importants de l'évaluation ;

— pour faire travailler l'ensemble des enseignants sur l'analyse de ces résultats qui ont été diffusés dans chaque établissement. Ce travail peut se faire au sein des équipes pédagogiques d'un établissement ou surtout dans des séances de formation : si l'objectif est maintenu et si le résultat n'est pas satisfaisant, la formation des enseignants devient un moyen d'action privilégié ;

— pour informer le public : les résultats détaillés sont publiés dans les brochures du ministère ; diverses réunions de travail ont eu lieu avec des journalistes spécialisés en éducation.

L'observatoire des Entrées dans la Vie active du CEREQ permet de même d'analyser l'évolution de telle ou telle filière de formation au cours du temps du point de vue des possibilités d'insertion dans la vie professionnelle.

Le pilotage d'une politique, avec inflexion en fonction des résultats de l'évaluation, peut également permettre une amélioration progressive de l'appréciation des mécanismes en jeu. Dans un tel contexte, la distinction entre évaluations *ex ante* et

17. Voir P. Mondon *L'évaluation du système éducatif* (EPP).

Encadré n° 3

UN EXEMPLE D'OBSERVATOIRE PERMANENT : LE DISPOSITIF D'ÉVALUATION DU MINISTÈRE DE L'ÉDUCATION NATIONALE

Le dispositif d'évaluation du ministère de l'Éducation nationale en matière pédagogique mis en place par le Service de la prévision, des statistiques et de l'évaluation a pour objet de permettre une évaluation des savoirs et savoir-faire des élèves aux niveaux clés de leur scolarité et en regard des objectifs fixés par les programmes pédagogiques.

Ceux-ci sont traduits en « objectifs opérationnels » susceptibles de faire l'objet d'épreuves qui en mesurent l'atteinte. Le travail d'élaboration des projets d'épreuves a associé des inspecteurs, des conseillers pédagogiques, des enseignants de différentes catégories, des chercheurs ; on a recherché un accord sur l'interprétation des objectifs contenus dans les programmes et sur le choix des épreuves. Les différents types de réponses sont identifiés par le biais de grilles de codage.

Des échantillons d'écoles et de collèges ont été tirés de façon aléatoire. Ces échantillons ont été stratifiés sur un petit nombre de variables (taille de l'établissement, structure de classes, environnement rural ou urbain), dont le contrôle permet de garantir la représentativité de l'échantillon sur l'ensemble des autres caractéristiques importantes (statut et ancienneté des enseignants, catégories socio-professionnelles des parents, etc.).

La taille des échantillons retenus est la suivante :

Ecoles		Collèges
CP	CM2	6 ^e
2 x 1 900 élèves	2 872 élèves	9 324 élèves
400 classes	492 classes	387 divisions 60 collèges

Le dispositif permet de suivre certains élèves au cours de leur scolarité car ce sont les mêmes établissements qui sont retenus pour les différents niveaux évalués. Les épreuves se sont ainsi déroulées :

	Juin 1979	Epreuves de fin de CP
Ecoles	Juin 1981	Epreuves de fin de CE2
	Juin 1983	Epreuves de fin de CM2

	Septembre 1980	Epreuves d'entrée en 6 ^e
Collèges	Juin 1982	Epreuves de fin de 5 ^e
	Juin 1984	Epreuves de fin de 3 ^e

La masse d'informations recueillies doit être utilisée de manière notamment à relier les acquisitions ainsi évaluées à l'environnement des élèves et ce, à partir d'analyses statistiques.

ex post s'atténue. L'évaluation devient partie intégrante de l'élaboration de la politique : on parle parfois d'« évaluation *ex tempore* »¹⁸.

Il convient toutefois d'observer que la dimension temporelle des effets d'une politique doit être prise en compte. En règle générale, l'observation O_n ne sera pas fonction de la seule politique X_{n-1} dont elle cherche à mesurer les conséquences. Elle résultera aussi des effets décalés des politiques précédemment mises en œuvre (X_{n-2} , X_{n-3} , etc.) qui poursuivaient certes les mêmes objectifs mais avec des méthodes peut-être différentes. Cette difficulté a été mise en évidence dans le cas du bilan du plan textile (encadré n° 11), rendu délicat par la succession des mesures antérieures qui ont affecté tout ou partie du secteur du textile-habillement et dont certaines peuvent continuer à produire des effets. On la rencontre dans toute politique réalisée à la marge d'un système sédimentaire.

Comme l'observation avant-après dont elle constitue une extension, le suivi de la mise en œuvre progressive de la politique peut répondre à des questions portant sur l'ampleur des modifications enregistrées. Elle permet plus difficilement de les attribuer à la politique plutôt qu'à des causes externes. L'absence d'un groupe témoin, susceptible d'aider à formuler des hypothèses sur les mécanismes en jeu, affaiblit en effet les conclusions qui peuvent être tirées de ce type de procédure d'évaluation. Toutefois, dans le cas où les séries temporelles sont suffisamment longues et correspondent à la fois à des périodes où la politique a été mise en œuvre et à des périodes où elle ne l'a pas été (« interrupted time series » dans la terminologie anglo-saxonne), il est parfois possible de tester des hypothèses de causalité.

4) Pour les plans de recherche relevant de l'avant-après strict ou du suivi de mise en œuvre, les techniques utilisées appartiennent à la catégorie (large) des analyses de séries temporelles, ou au recours à des données de panel.

Les **analyses de séries temporelles** constituent la catégorie de techniques la plus fréquemment utilisée dans l'évaluation avant-après, avec suivi ou non. Ces analyses cherchent à tirer parti des ruptures d'évolution des séries étudiées, correspondant aux périodes pendant lesquelles la politique produit ses effets. Elles ne nécessitent pas forcément un dispositif d'observation préalable puisque certaines informations peuvent être reconstituées à partir de données historiques. En outre, on l'a dit, leur transcription graphique présente l'avantage d'être facilement compréhensible.

Les analyses de séries temporelles présentent toutefois un certain nombre de difficultés méthodologiques répertoriées dans la littérature (voir par exemple D. T. Campbell, *Assessing...*) :

— en raison de problèmes d'autocorrélation, l'estimation par la méthode des moindres carrés ordinaire est en général inapplicable. Les méthodes du type Box-Jenkins apparaissent plus satisfaisantes, bien qu'elles nécessitent un nombre suffisant d'observations. Certains développements de ces méthodes concernant « l'analyse des interventions »¹⁹ permettent notamment de tenir compte de

18. J. P. Nioche *De l'évaluation à l'analyse des politiques publiques* (1982 — Revue française de Sciences politiques, vol. 32, n° 1).

19. Box, G. C. Tiao *Intervention analysis with applications to economic and environmental problems*, 1975, Journal of the American Statistical Association, n° 349.

variables indicatrices caractérisant les périodes de mise en œuvre de la politique : elles apparaissent donc particulièrement adaptées dans une perspective d'évaluation ;

— la tendance à modifier le système de collecte de données à l'occasion de la mise en œuvre d'une nouvelle politique doit être autant que possible évitée car elle est susceptible de rendre ininterprétable la modification de certains indicateurs ;

— dans le cas d'une politique destinée à répondre à la brusque dégradation d'un indicateur, les améliorations dues à la politique risquent de ne pas pouvoir être distinguées de corrections spontanées provenant de certaines forces de rappel. Dans un même ordre d'idées, on a mentionné précédemment les « perversions » possibles des indicateurs ;

— les effets de politiques graduelles sont en général indétectables par ces méthodes, ce qui limite notablement leur portée. Ce type de difficulté a été rencontré à l'occasion de l'évaluation du plan textile : l'ampleur des « cycles » affectant ce secteur rend malaisé la détection des ruptures de tendance associées au plan ;

— la nécessité de disposer de séries suffisamment longues exige fréquemment de recourir à des indicateurs mis au point pour un autre propos et qui peuvent de ce fait se révéler peu pertinents pour l'évaluation que l'on cherche à réaliser.

Les analyses s'appuyant sur l'**observation de panels** sont celles qui présentent les meilleures garanties de validité en l'absence de groupe témoin. Il en est, en particulier, ainsi des estimations économétriques mêlant coupes instantanées et séries temporelles. Toutefois les méthodes économétriques à utiliser dans un tel contexte soulèvent des difficultés techniques qui peuvent rendre leur mise en œuvre difficile : erreur sur la liste des variables explicatives, modèles à équations simultanées, modèles à erreurs composées²⁰. Elles nécessitent des délais notables et requièrent des compétences techniques spécifiques.

II - 3 L'évaluation par comparaison avec un groupe témoin non équivalent

Les plans de recherche correspondants reposent sur la comparaison du groupe cible, affecté par la politique, avec un groupe témoin, mais dont les caractéristiques ne sont pas nécessairement (statistiquement) semblables à celles du groupe cible. Compte tenu d'un croisement avec observation « après seulement » ou observation « avant-après » on aboutit aux deux schémas représentatifs suivants (schémas 4 et 5).

20. Sur ce point voir le n° spécial des annales de l'INSEE sur l'économétrie des données de panel (1978, n° 30-31).

Groupe cible :	X	O

Groupe témoin non équivalent :		O'

Schéma 4 - Comparaison avec un groupe témoin non équivalent : après seulement

Groupe cible :	O ₁	X	O ₂

Groupe témoin non équivalent :	O' ₁		O' ₂

Schéma 5 - Comparaison avec un groupe témoin non équivalent : avant-après

1) Les groupes témoins non équivalents peuvent provenir d'un acte explicite de l'évaluateur : on parlera dans ce cas de **quasi-expérimentation** (la restriction du « quasi » faisant évidemment référence au caractère non équivalent du groupe). Ils peuvent également résulter d'une **observation** « spontanée » : dans ce cas, l'évaluateur n'intervient pas dans la constitution des groupes comparés, ni dans le « traitement » (pour employer un terme emprunté à l'évaluation des thérapeutiques) qui leur est appliqué.

Les méthodes de **quasi-expérimentation** permettent à l'évaluateur d'avoir une certaine marge de manœuvre dans la constitution des groupes. Elles ont fait l'objet de plusieurs applications en santé publique (détection précoce du cancer du sein en Grande-Bretagne). Les tentatives de mise en œuvre pour des politiques économiques et sociales sont moins nombreuses (divers programmes socio-économiques aux États-Unis).

La comparaison à des groupes témoins non équivalents, simplement « observés », est une pratique plus courante, notamment en France.

Une raison en est que l'évaluateur n'a pas à intervenir dans la mise en œuvre de la politique. Il met simplement à profit le fait que cette politique n'est pas l'objet d'une application universelle ou uniforme. D'où des difficultés moindres d'insertion administrative et un coût plus bas que dans le cas d'une démarche quasi-expérimentale.

Les exemples présentés au groupe ont été assez nombreux et portent sur divers secteurs :

- évaluation du plan textile : comparaisons internationales avec l'évolution des secteurs textiles belges et italiens ; comparaison de l'évolution du secteur textile avec l'évolution du secteur (relativement similaire) du cuir-chaussure ;
- comparaisons internationales (avec le Danemark) et régionales (entre plusieurs départements français) de la prophylaxie de la brucellose, évaluée dans le cadre de l'étude prioritaire RCB sur les aides à l'élevage bovin (encadré n° 4) ;
- comparaison entre la Bretagne et les Pays-Bas en matière de développement de la filière porc ;

— comparaison entreprises aidées-entreprises non aidées en matière d'aides négociées à l'emploi (étude RCB sur les aides à l'emploi) ;

— évaluation de l'effet de l'allocation de parent isolé sur l'offre de travail (encadré n° 10)²¹.

La simple observation de groupes non équivalents est parfois jugée de pertinence limitée pour déduire des relations causales, dans la mesure où l'on n'est pas, moins encore que dans le cas d'une quasi-expérimentation bien contrôlée, assuré de la comparabilité des groupes (ces insuffisances ont été particulièrement mises en évidence dans le domaine de la santé publique²²).

Ce jugement réservé mérite d'être nuancé dans le cas de politiques publiques de nature incitative, et qui ont donc vocation à être proposées, et non imposées, à l'ensemble des agents potentiellement concernés. C'est le cas en particulier de l'ensemble des systèmes d'aides qui entraînent en contrepartie certaines contraintes (l'aide proposée dans le cadre du plan textile s'accompagnait ainsi d'obligations en matière d'investissement et d'emploi). Dans ce cas, le partage en deux groupes (agents aidés, agents non aidés) est révélateur de mécanismes de choix qui font partie intégrante de la politique suivie et donc de son évaluation. La comparaison aidés-non aidés devrait notamment permettre de comprendre pourquoi un agent économique décide d'adhérer ou de ne pas adhérer au système proposé, élément important de la comparaison entre procédures d'aides.

Pour conclure cette discussion sur les mérites et limites respectifs de la quasi-expérimentation et de l'observation, on notera qu'il existe des situations intermédiaires. Ainsi, l'évaluation de la prévention de la carie dentaire par adjonction du fluor au sel alimentaire en Suisse repose sur la comparaison de groupes issus de l'observation mais dont les éléments de comparabilité ont pu être précisés appréciés : l'observation est dans ce cas « raffinée » et rejoint la quasi-expérimentation.

2) L'étude *ex post* avec des groupes non équivalents est fréquemment utilisée, notamment dans le cadre de quasi-expérimentations, lorsque la politique est mise en place (schéma 4, comparaison avec un groupe témoin non équivalent : après seulement) avant le dispositif d'évaluation. Dans ce cas on cherche à comparer *ex post* un groupe ayant subi la politique et un groupe ne l'ayant pas subie. Les deux groupes sont non équivalents en ce sens qu'ils ne résultent pas d'un tirage au sort.

Un tel plan de recherche ne signifie pas nécessairement que l'on ne dispose pas d'informations antérieures à la mise en œuvre de la politique (ces informations peuvent souvent être reconstituées ou collectées *ex post* auprès des groupes observés) : il correspond seulement à la non existence d'observations réalisées *ex ante* sur une même échelle et de même nature que celles disponibles *ex post*.

Le fait que les groupes ne soient pas équivalents est susceptible d'affaiblir sensiblement les conclusions en terme de relations de causalité, par rapport à ce qu'il serait possible de déduire, dans certains cas, d'une expérimentation avec tirage au sort. Il n'est en effet pas possible d'éliminer des facteurs autres que la

21. Voir J.-C. Ray, *Mères isolées, API et travail* (1984, communication aux IV^{es} Journées d'économie sociale, ronéo).

22. Voir les exemples donnés par D. Schwartz (in EPP), en particulier celui relatif à la prévention du cancer du col de l'utérus.

COMPARAISON DE GROUPES NON ÉQUIVALENTS ISSUS DE
L'OBSERVATION :
L'EXEMPLE DE L'ÉVALUATION DE LA POLITIQUE DE
PROPHYLAXIE DE LA BRUCELLOSE

La politique de prophylaxie de la brucellose a réellement commencé en France en 1964 à la suite d'une directive de la Communauté économique européenne visant à l'éradication de cette maladie.

Jusqu'en 1972, la politique suivie a été celle de la prophylaxie sanitaire (abattage), suivant en cela l'exemple danois (le Danemark ayant pratiquement éradiqué la maladie en adoptant cette politique). Après 1973, on s'est orienté vers une politique de prophylaxie médicale (vaccination et traitement). La politique suivie en France a donc été très régionalisée, et même départementalisée, suivant la période durant laquelle l'effort d'éradication de la brucellose a été mené.

Ainsi l'étude sur les aides à l'élevage bovin utilise une comparaison entre 4 départements dont les cheptels sont numériquement voisins mais qui ont mené des politiques différentes :

- prophylaxie sanitaire en Ille-et-Vilaine et Morbihan ;
- prophylaxie médicale en Vendée ;
- prophylaxie médicale puis sanitaire en Saône-et-Loire.

On a cherché à vérifier si une politique plus systématique de prophylaxie médicale n'aurait pas été préférable en France, comme le suggère par exemple la comparaison des résultats obtenus en Vendée et en Ille-et-Vilaine. L'échec de la transposition de l'exemple danois peut s'expliquer par le taux d'infection beaucoup plus élevé en France (de l'ordre de 40 %) : la brucellose cesse alors d'être une maladie de l'animal pour devenir une maladie du troupeau. De plus, l'effort d'épidémiologie avait été réalisé sur plus longue période au Danemark, qui présente en outre l'avantage d'un plus grand isolement (presqu'île).

politique suivie, pouvant constituer des causes d'écart entre les observations relatives aux divers groupes. Il est évidemment recommandé de chercher à constituer des groupes aussi homogènes que possible mais ceci peut être difficile à atteindre comme le montre l'exemple du bilan du plan textile :

— soit parce que les groupes diffèrent par certaines de leurs caractéristiques : ainsi les secteurs du textile-habillement et du cuir-chaussure, comparés dans l'évaluation du plan textile, sont affectés par des effets de mode différents et ne se caractérisent pas par les mêmes structures de marchés. En outre, l'évaluateur ne pourra pas toujours, *ex post*, identifier toutes les différences pertinentes ;

— soit parce que l'information n'est pas d'une qualité comparable pour les deux groupes : la comparaison entre les secteurs textiles français et étrangers s'est ainsi heurtée à des difficultés pour obtenir des données concernant les pays voisins ;

— soit parce que le groupe de contrôle a lui-même été affecté par une autre politique dont les effets n'ont pas été complètement évalués. On se trouve alors dans une situation décrite par le schéma 4 bis.

Groupe cible :	X	O
Groupe témoin non équivalent :	X'	O'

Schéma 4 bis - Évaluation par comparaison avec un groupe témoin non équivalent : après seulement et présence d'une politique perturbatrice.

Un exemple de ce dernier schéma est fourni par la comparaison du secteur du textile-habillement à celui du cuir-chaussure, qui avait lui-même bénéficié d'un plan d'aide dont les effets pouvaient continuer à se manifester lors du bilan du plan textile.

Ces diverses difficultés expliquent les réticences à l'égard de ce type de plans de recherche exprimées par les praticiens de l'expérimentation avec tirage au sort. Quelles que soient les précautions prises pour constituer le (ou les) groupe(s) de contrôle, il n'est guère possible d'éviter qu'ils présentent des différences. Rien n'indique *a priori* qu'elles aient des effets négligeables par rapport à ceux de la politique que l'on évalue. C'est ainsi que dans le domaine de la santé, certaines comparaisons fondées sur l'observation d'hôpitaux différents ont été jugées non pertinentes²³.

Il existe des moyens d'alléger ces difficultés, en **diversifiant les groupes de contrôle**, lorsque ceci est possible, ou en recourant à des **techniques d'analyse multivariée**²⁴ pour estimer l'effet des facteurs parasites.

23. Exemple cité par D. Schwartz.

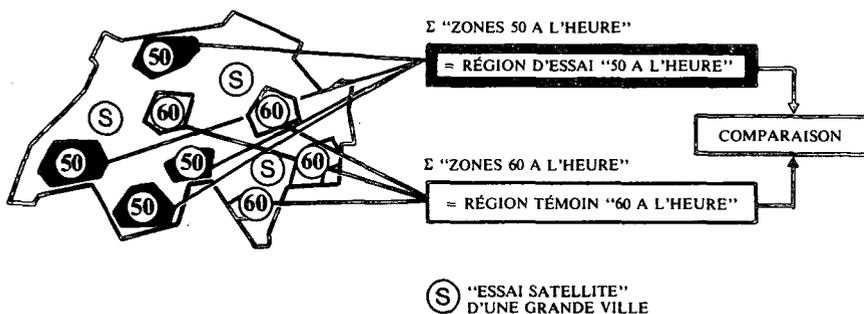
24. L'analyse multivariée a notamment été utilisée pour l'insertion des 16-18 ans (cf. encadré n° 6). Les conclusions auxquelles elle a conduit dans ce cas sont toutefois controversées, d'une part parce que les groupes ne sont pas réellement équivalents et d'autre part en raison du choix des variables explicatives retenues.

ESSAI « 50 A L'HEURE » (SUISSE)

Objectif : Déterminer les gains de sécurité résultant d'un abaissement de la vitesse limite des véhicules dans les agglomérations. Au préalable existaient de grandes divergences, soit dans les avis exprimés sur l'efficacité d'une telle mesure, soit dans l'interprétation des données disponibles.

Modalités : L'opération a été lancée en novembre 1978 par une ordonnance du Conseil fédéral suivie le 21 avril 1980 d'une ordonnance du département fédéral de la Justice et de la Police chargé de la conduite globale des travaux. Le rapport final a été déposé en mars 1983. Des gouvernements cantonaux et des villes ont participé à l'opération. Celle-ci a associé de nombreux instituts et prestataires spécialisés.

Méthode : La méthode retenue pour l'essai a consisté à faire ressortir les caractéristiques considérées comme importantes en analysant de manière détaillée les accidents, en questionnant les usagers de la route pour connaître leur attitude, en observant leur comportement dans le trafic et en effectuant des mesures. Les données ainsi obtenues ont été ensuite contrôlées, dans le sens d'« hypothèses », au moyen de techniques statistiques. Lors de ces enquêtes, ce sont les procédures de comparaison avant/après et avec/sans qui ont été appliquées. Cela a supposé la constitution de deux régions d'enquête semblables, une région d'essai 50 et une région témoin 60, composées chacune de diverses zones et représentatives de l'ensemble des zones urbaines du territoire.



Difficultés : Un recours judiciaire a attaqué comme non valable l'abaissement à titre expérimental de la vitesse maximale autorisée dans des localités. Il s'agissait de contester la punition infligée à un automobiliste qui avait dépassé le 50 km/h. Le tribunal fédéral n'a pas suivi cette voie.

Conclusions : Les effets de l'abaissement de la vitesse limitée sont globalement favorables en ce qui concerne le nombre et la gravité des accidents. Toutefois la sécurité des véhicules à deux roues ne progresse pas et les points noirs ne sont pas vraiment éliminés. Par ailleurs, les enquêtes ont confirmé que 50 km/h constitue la borne inférieure des limites possibles.

Ce type d'évaluation « par répétition » permet d'arriver à des conclusions assez solides lorsque les facteurs autres que la politique proprement dite peuvent être mis en évidence. Ainsi l'évaluation de la prophylaxie de la brucellose a permis de comprendre pourquoi le modèle danois (prophylaxie sanitaire par abattage des bêtes malades) n'était pas transposable en France où une prophylaxie médicale par vaccination et soins s'est révélée plus efficace : d'une part, le Danemark est une presque île (ce qui limite le risque de contagion), d'autre part le pourcentage du troupeau atteint par la maladie s'est avéré beaucoup plus élevé en France qu'au Danemark (la brucellose cesse alors d'être une maladie de l'animal pour devenir une maladie du troupeau) (voir encadré n° 4). La multiplication des groupes (le Danemark et 4 départements français ayant subi des politiques très différentes) explique en partie la robustesse de ses conclusions. Rappelons néanmoins que la répétition de comparaisons sur une multiplicité de groupes n'a pas valeur de confirmation à tout coup : dans certains cas, les répétitions peuvent comporter toutes le même biais et la confirmation est alors une duperie. Pour faire en sorte que la confirmation soit véritable, il faut répéter le plan de recherche dans des situations différentes, excluant l'existence d'un tel biais systématique. A cet égard, **il n'est pas exclu que multiplier progressivement les groupes de contrôle non équivalents se révèle, à degré de certitude égale, aussi coûteux que de construire d'emblée un dispositif expérimental « propre ».**

4) L'évaluation par observation avant et après avec groupe de contrôle non équivalent (schéma 5) suppose l'existence d'un dispositif d'observation *ad hoc* préalable à la politique.

Dans ce cas, la constitution des groupes par quasi-expérimentation et le recours à une estimation statistique de modèles microéconomiques mêlant coupes instantanées et séries temporelles sont envisageables. Pour tester des relations de causalités, cette procédure d'évaluation offre les garanties les meilleures à l'exception de l'expérimentation pure (lorsque celle-ci est possible et n'entraîne pas de phénomènes pervers). Elle assure une certaine validité aussi bien externe qu'interne à l'évaluation.

En contrepartie, elle peut se révéler coûteuse en raison de la lourdeur des dispositifs impliqués et de mise en œuvre délicate du point de vue des méthodes statistiques.

5) pour ces divers plans de recherche avec groupes non équivalents, après seulement ou avant-après (schémas 4, 5, 4 bis), il est fréquent de recourir à une simple analyse d'indicateurs mesurant l'effet de la politique. Ce type de comparaison est facile à mettre en œuvre mais il ne permet pas toujours d'aboutir à des conclusions très robustes. Là encore, le caractère non équivalent des groupes considérés ne permet pas d'écarter l'influence de facteurs parasites.

Des tests d'analyse de la variance ou de la covariance sont susceptibles d'améliorer la fiabilité des résultats sur les effets de la politique. L'estimation économétrique sur coupes transversales (observations *ex post*) ou sur panels (observations *ex ante* et *ex post*) peut mettre en évidence le poids respectif des différentes variables explicatives (y compris celles caractérisant la politique) une fois celles-ci identifiées et quantifiées. On retrouve bien sûr certaines difficultés techniques, déjà présentes pour les comparaisons sans groupe témoin : erreur de spécification sur la liste des

L'ÉVALUATION DU DISPOSITIF DE FORMATION ET D'INSERTION DESTINÉ AUX JEUNES DE 16-18 ANS

Ce dispositif, mis en place en 1982, avait pour objet de fournir une qualification professionnelle aux jeunes sortis sans formation du système scolaire et de faciliter leur insertion sociale. Les stages du dispositif étaient de trois types : stages d'orientation (d'une durée brève), stages d'insertion et stages de qualification (de six mois à deux ans).

Au cours de l'année 1982-1983, 115 000 jeunes ont été accueillis par les 874 permanences d'accueil, d'information et d'orientation et les 62 missions locales : 105 000 d'entre eux ont été orientés vers des stages du dispositif.

Le dispositif avait dès l'origine fait l'objet d'un suivi par une mission d'observation et d'évaluation confiée au Centre de formation de formateurs du Conservatoire national des arts et métiers, sous la responsabilité de G. Malglaive et à 15 centres de recherche universitaires régionaux. Dans ce cadre, l'Institut de recherche sur l'économie de l'éducation de l'université de Dijon a par exemple réalisé une évaluation de ce dispositif à partir d'une enquête réalisée en février 1984 auprès de jeunes passés fin 1982 par trois structures d'accueil situées dans la région dijonnaise (une en milieu urbain, deux en milieu rural)*.

La population de l'enquête comprend 258 individus répartis ainsi :

	Dispositif	Hors dispositif	Ensemble
Urbains	119	37	156
Ruraux	72	30	102
Ensemble	191	67	258

Les caractéristiques individuelles des individus ayant bénéficié d'un stage et des autres (sexe, nationalité du père, taille de la famille, situation professionnelle des parents) ont été comparées au sein des sous-échantillons « urbain » d'une part, « rural » d'autre part. Les jeunes du dispositif et les jeunes hors-dispositif n'apparaissent pas significativement différents à l'intérieur de chacun de ces sous-échantillons. Par contre des différences marquées séparent jeunes urbains et jeunes ruraux.

A partir d'estimations économétriques de modèles d'accès à l'emploi et de détermination du salaire perçu par l'ensemble des individus employés à la date de l'enquête, les auteurs concluent à l'absence d'impact du dispositif : d'autres variables (sexe, nationalité du père) jouent par contre un rôle significatif. Le dispositif aurait donc principalement un effet de rétention de l'offre de travail. Le fait d'avoir suivi un stage s'est traduit en outre, à court terme, par un manque à gagner pour les jeunes concernés.

Ces conclusions ont été contestées par certains spécialistes en raison d'une part de la non comparabilité des groupes, d'autre part du choix des variables explicatives retenues : ceci illustre bien les difficultés et les limites de l'analyse multivariée lorsqu'il n'existe pas de théorie structurée susceptible d'appréhender les mécanismes de la politique à évaluer.

* Des éléments de cette étude ont été présentés en septembre 1984 aux journées d'évaluation organisées par l'Association d'économie sociale et le Commissariat général du plan. Cf. J.-J. Paul : éléments d'évaluation du dispositif de formation et d'insertion destiné aux jeunes de 16-18 ans.

variables explicatives, modèles à erreurs comparées (dans le cas de l'économétrie sur panels) etc.

Malgré ses vertus, ce type d'approche est encore assez peu développé, notamment en raison des compétences techniques qu'il exige. L'évaluation des conséquences de l'allocation de parent isolé en termes de désincitation au travail, réalisée par J.-C. Ray, en fournit une illustration (cf. encadré n° 10) : un modèle microéconomique a été construit, qui explique la probabilité qu'une femme isolée travaille en fonction des revenus salariaux et non salariaux dont elle dispose, et d'autres variables comme le niveau d'éducation et le nombre d'enfants. Ce modèle a été estimé sur deux échantillons, l'un constitué de femmes isolées bénéficiant de l'API, l'autre de femmes isolées n'ayant plus droit à cette prestation. Il a ainsi été possible de quantifier l'impact de l'API et de simuler les conséquences de sa suppression ou d'une modification de ses conditions de versement.

La modélisation (au sens large) est un point de passage obligé pour déterminer l'influence des facteurs externes à la politique : elle peut donc constituer un instrument très utile pour améliorer la validité de l'évaluation. Ceci est particulièrement vrai lorsque l'on dispose de données microéconomiques comme c'est le cas si l'on compare des groupes affectés et des groupes non affectés par la politique.

II - 4 L'évaluation par comparaison avec un groupe témoin équivalent ou expérimentation vraie

Les plans de recherche correspondants reposent sur la comparaison du groupe cible, affecté par la politique, avec un groupe témoin « équivalent », c'est-à-dire dont les caractéristiques sont (statistiquement) semblables à celles du groupe cible. La propriété d'équivalence est évidemment fondamentale pour la validité du plan de recherche, d'où l'importance de la procédure de constitution du groupe témoin.

L'évaluation par comparaison avec groupe témoin équivalent est représentée par le schéma 6.

Groupe cible (ou échantillon représentatif)	:	X	O
Groupe témoin équivalent	:		O'

Schéma 6 - Comparaison avec un groupe témoin équivalent

Comme on l'a dit, le point crucial pour la validité interne de l'évaluation est l'équivalence des deux groupes.

La seule façon de la garantir est de les constituer par tirage au sort.

Il existe une littérature technique abondante sur la constitution d'échantillons équivalents. Historiquement, elle s'est développée à partir de travaux expérimentaux en agriculture et en évaluation de thérapeutiques.

Dans ce dernier domaine, l'expérimentation vraie apparaît comme la seule procédure rigoureuse d'évaluation : D. Schwartz (voir EPP) montre ainsi que les évaluations quasi-expérimentales ou issues d'observations ne présentent pas une validité incontestable, même lorsque l'on est très attentif à la comparabilité des groupes traités et témoins et qu'on utilise des méthodes multivariées pour éliminer les facteurs parasites liées à la non-équivalence elle-même. Ainsi l'étude de l'effet du fluor sur la carie dentaire, en Suisse, a été menée en comparant des individus provenant de cantons voisins et démographiquement semblables mais n'appliquant pas la même politique sanitaire. Malgré plusieurs précautions de type expérimental (mesures standardisées et à l'aveugle), les différences entre politiques sanitaires affectant les différents groupes limitent leur comparabilité et donc la portée des conclusions. De manière plus générale, les procédures de quasi-expérimentation retiennent souvent les groupes traités et témoins en fonction de contraintes de faisabilité (ce fut par exemple le cas dans l'essai de détection du cancer du sein en Grande-Bretagne), ce qui là encore peut conduire à la non-équivalence de ces groupes malgré toutes les précautions prises et à une non conclusivité de l'étude. On retrouve ce fait que, dans bien des cas, l'expérimentation vraie n'est chère qu'avant l'évaluation.

Les insuffisances des méthodes de comparaison de groupes non équivalents observés en évaluation de thérapeutiques se retrouvent évidemment en matière d'évaluation de politiques. **Faut-il en conclure que l'expérimentation avec tirage au sort doit être systématiquement privilégiée pour évaluer les politiques publiques ? En dépit d'exemples significatifs de l'apport de ce plan de recherche, notamment aux États-Unis, il n'est pas réaliste d'aller aussi loin.** En effet, l'application de méthodes expérimentales à l'évaluation de politiques n'est pas exempte de difficultés, qu'il s'agisse d'expérimentation pure ou de quasi-expérimentation :

— on ne peut pas toujours être assuré que les **comportements des agents** seront les mêmes dans un contexte de généralisation de la politique et dans un contexte expérimental. Campbell mentionne cette difficulté (« **Hawthorne effect** ») dans l'expérimentation de l'impôt négatif du New Jersey (encadré n° 7). Les familles « expérimentales » se savaient seules à bénéficier de ce programme, avec une attention particulière du public et des médias : elles eurent donc tendance à « montrer le bon exemple » en continuant à travailler comme avant. Le phénomène était accentué par une réalisation d'échantillonnage aléatoire plus en fonction des caractéristiques des familles que des zones : les bénéficiaires de l'impôt négatif étaient donc entourés de ménages aussi pauvres mais qui ne percevaient pas cette allocation. Ces facteurs de voisinage ont pu infléchir leur attitude, par rapport à un contexte d'application généralisé du régime fiscal soumis à évaluation expérimentale ;

— le fait que l'expérimentation soit limitée dans le temps, alors que la politique peut avoir, une fois généralisée, vocation à être permanente, est également susceptible d'introduire un **biais**. Dans le cas de l'impôt négatif du New Jersey, les participants savaient que l'expérimentation durerait seulement trois ans : la

prudence les incitait donc à conserver leur emploi de manière à retrouver ensuite la même situation financière qu'auparavant ;

— le phénomène de lassitude joue également dans les expérimentations de longue durée. Toujours dans le cas de l'expérimentation de l'impôt négatif du New Jersey, le taux de défection fut de 25 % dans le groupe témoin mais seulement de 6,5 % dans le groupe expérimental ayant reçu la rémunération la plus élevée (cf. Campbell) : ces différences sont susceptibles de produire des biais dans l'analyse des résultats. Ce phénomène de lassitude et le refus de participer à l'expérimentation tendent concrètement à rapprocher certaines expérimentations pures des quasi-expérimentations ;

— il existe souvent des **interactions entre les groupes** qui peuvent fausser les résultats de l'évaluation. Supposons, par exemple, que l'on répartisse, dans le cadre d'une expérimentation, les entreprises d'un secteur en deux groupes : celles qui bénéficieront d'un système d'aide dont on cherche à faire l'évaluation, et celles qui n'en bénéficieront pas. Par simple phénomène de « cannibalisme », on risque d'enregistrer une redistribution des parts de marché au profit du premier groupe, avec une augmentation de l'emploi et des investissements chez les entreprises aidées et une dégradation chez les autres. Il serait erroné de déduire, dans ce cas et pour un marché limité, que le système aurait des conséquences globales favorables une fois généralisé ;

— la nécessité de prendre en compte **des effets à long terme** qui peuvent différer sensiblement des effets enregistrés à court terme constitue une difficulté pour toutes les procédures d'évaluation. Cette difficulté se trouve renforcée lorsqu'il s'agit de mettre en place un dispositif coûteux d'expérimentation. En outre l'expérimentation est destinée à se dérouler sur une durée brève, au terme de laquelle, les effets de la politique étant analysés, l'on décidera si celle-ci doit effectivement être mise en œuvre. **L'expérimentation ne peut guère participer à l'élaboration progressive d'une politique sauf à ne mesurer à chaque fois que les effets à court terme des modifications introduites ;**

— l'expérimentation peut poser des problèmes de compatibilité avec les **contraintes juridiques** d'égalité de traitement.

Dans le domaine de l'évaluation des thérapeutiques, on a pu dégager certaines parades à ces difficultés :

— la première repose sur le recours à des essais dits « pragmatiques », par opposition aux essais explicatifs. Les premiers visent la décision, les seconds la connaissance. Le protocole d'un essai pragmatique doit assurer une identité entre les conditions de l'essai (de traitements, de patients, de critères de jugement) et celles qui prévaudront si l'on généralise la thérapeutique. L'analyse des résultats est également différente de celle de l'essai explicatif²⁵ ;

— la seconde vise à éviter les effets de « cannibalisme » ou de contagion. Elle consiste à prendre comme unité de base, non le sujet, mais un groupe de sujets.

25. Le principe de ces essais est décrit dans l'ouvrage de D. Schwartz, R. Flamant et J. Lellouch, *L'essai thérapeutique chez l'homme*, (Flammarion).

En dépit des difficultés qu'elle soulève et pour lesquelles il existe ces parades, l'expérimentation pure est incontestablement supérieure aux autres plans de recherche en matière de validité interne. Sa supériorité en matière de validité externe et conceptuelle a pu être, par contre, discutée dans certains cas : Campbell (« assessing » — op. cit.) note ainsi que l'évaluation de l'impôt négatif aurait une validité interne moindre mais des validités conceptuelle et externe supérieures si la politique était appliquée de manière globale et si l'on s'appuyait, pour juger de ses effets, sur des analyses de séries temporelles et des comparaisons avec des villes canadiennes. Il reste que l'expérimentation vraie est particulièrement adaptée lorsque lever une ambiguïté représente un enjeu important (c'est le cas par exemple si l'on veut tester l'efficacité d'un vaccin).

Les techniques utilisables dans le cadre de l'expérimentation vraie sont évidemment semblables à celles mentionnées pour les comparaisons avec groupes témoins non équivalents. Le recours à la modélisation, compte tenu des informations acquises au travers de l'expérimentation, paraît particulièrement adapté pour élargir la validité conceptuelle et externe de l'évaluation²⁶.

5 Évaluation et théorie : de l'étude de cas à la modélisation explicative

L'aptitude à quantifier les effets d'une politique est un critère de classification commode des techniques d'évaluation. De ce point de vue, l'étude de cas et la modélisation constituent probablement des exemples extrêmes. Entre ces extrêmes, la plupart des plans de recherche font appel à des combinaisons de techniques pouvant aller de l'interprétation qualitative d'enquêtes statistiques jusqu'à l'utilisation de techniques statistiques élaborées (analyse de séries temporelles, analyse multivariée, construction de typologies, etc.).

1) Même si les autres techniques ne peuvent pas être mises en œuvre, il reste toujours à l'évaluateur la possibilité de réaliser des monographies. Celles-ci ne permettent pas d'avoir une appréciation exhaustive ou statistiquement non biaisée des effets d'une politique. En revanche, elles peuvent fournir des indications précieuses sur la manière dont cette politique est mise en œuvre et sur ses répercussions concrètes.

Il en est ainsi de l'évaluation de l'impact des ordonnances de 1982 sur les 39 heures et la 5^e semaine de congés payés²⁷ (encadré n° 8). Ce type d'approche est donc très utile lorsque les idées que l'on peut avoir *a priori* sur le fonctionnement de la politique sont particulièrement vagues (cas, par exemple, de nombre de politiques sociales).

26. On trouvera dans l'ouvrage de Killingsworth *Labor Supply* (1983, Cambridge University Press) le compte rendu de travaux économétriques sur le thème « offre de travail, taxation et transferts » et recourant à l'estimation de comportements microéconomiques sur données générées par l'expérimentation.

27. Cf. M. Pépin et D. Tonneau — op. cit.

L'EXPÉRIMENTATION D'UN IMPÔT NÉGATIF DANS LE NEW JERSEY*

Cette expérimentation, réalisée entre 1969 et 1972, a concerné 1216 familles ayant un chef de famille masculin. Son objectif était d'une part d'analyser les modifications de l'incitation à travailler entraînées par l'application de l'impôt négatif, d'autre part de déterminer dans quelle mesure cette allocation serait utilisée par les familles bénéficiaires pour améliorer leurs compétences professionnelles.

Parmi les 1216 familles ayant participé à l'expérimentation, 725 familles (groupe expérimental) ont reçu un impôt négatif portant leur revenu à un niveau compris entre 3 000 et 4 000 \$ par an pour une famille de 4 personnes. Le groupe témoin était constitué des 491 autres familles qui n'ont reçu aucune allocation de cette sorte, mais qui ont collaboré aux entretiens mensuels.

L'expérimentation a duré 3 ans et a coûté 8 millions \$ dont 3 millions \$ au titre des dépenses administratives et des paiements versés aux participants, et 5 millions \$ au titre des dépenses de recherche.

Huit systèmes d'impôt négatif différents suivant le montant de l'allocation versée et l'incitation à augmenter ses propres ressources ont été testés. Le tableau ci-dessous donne le niveau de base du revenu minimum garanti (c'est-à-dire le montant que percevrait une famille sans autre source de revenus) et le taux de taxation implicitement retenu pour les revenus dépassant ce niveau de base :

Système d'impôt négatif	Revenu garanti (en pourcentage du seuil de pauvreté)	Taux implicite de taxation (%)
1	50	30
2	50	50
3	75	30
4	75	50
5	75	70
6	100	50
7	100	70
8	125	50
Groupe témoin	0	0

* Voir notamment — David N. Kershaw : *the New Jersey negative income tax experiment ; a summary of the design, operations and results of the first large-scale social science experiment*. Extrait de : *Social Research and Public Policies* - G.M. Lyons Ed. The Dartmouth OECD Conference. Hanover : Dartmouth 1975. — Killingsworth : *Labour supply, Cambridge Survey of economic literature*. Cambridge University Press, 1983.

Encadré n° 7 (suite)

Les effets de l'impôt négatif sur l'offre de travail ont été estimés économétriquement à partir d'une relation faisant intervenir comme autres variables explicatives l'âge du chef de famille, son niveau d'instruction, le nombre d'adultes dans la famille, le nombre et l'âge des enfants, la localisation, les revenus et l'offre de travail de la famille avant l'expérimentation.

On a constaté que le taux d'activité des hommes chefs de famille a été très peu diminué par la perception de l'impôt négatif (- 1,5 %). Il en est de même pour le nombre d'heures travaillées (- 2 %).

Le fléchissement du taux d'activité et du temps de travail de leurs épouses a été nettement plus marqué (- 23 %). L'impôt négatif a en effet constitué pour elles une incitation à se consacrer davantage aux tâches domestiques (il y a en moyenne 4 enfants par familles) et à renoncer à un emploi peu rémunéré. Toutefois, le taux d'activité moyen des femmes était faible au départ (4,4 heures par semaine), la diminution absolue du temps de travail est donc restée modeste.

Si l'on considère les familles dans leur ensemble, le nombre de personnes employées par famille a été inférieur de 9,5 % dans le groupe expérimental à ce qu'il était dans le groupe de contrôle. Le nombre d'heures travaillées a été inférieur à 8,7 %. Cet effet provient essentiellement de la plus grande proportion de femmes restant à la maison et de jeunes terminant leurs études secondaires.

Si, dans l'ensemble, les conclusions de l'expérimentation du New Jersey ont été conformes à ce qu'attendaient les économistes, certains doutes ont été émis sur leur interprétation, notamment en raison des modifications qui ont affecté, en cours d'expérimentation, la législation sociale de cet État.

Trois autres expérimentations d'impôt négatif ont été mises en œuvre afin de corroborer les résultats obtenus dans le New Jersey :

— la première a porté, à partir de 1970, sur 800 familles à faible revenu dans des zones rurales d'Iowa et de Caroline du Nord. Elle a donné des résultats similaires à ceux du New Jersey ;

— l'expérience de Gary (Indiana), réalisée de 1971 à 1974, concernait 1 800 familles dont 60 % dirigées par des femmes. Elle a conduit à un résultat paradoxal : le remplacement de l'allocation pour enfant à charge par un impôt négatif conduisant à des prestations plus élevées, ne s'est pas traduit par une désincitation au travail ;

— l'expérimentation de Seattle et de Denver, réalisée sur une plus grande échelle (4 800 familles) semble conduire à un impact de l'impôt négatif sur l'offre de travail plus marqué que les précédentes. Elle a en outre permis aux chercheurs d'élaborer un modèle de simulation.

L'étude de cas est indispensable au niveau exploratoire ou lorsque la complexité des effets de la politique ne permet pas d'adopter une démarche analytique valide. Un exemple de l'apport de l'étude de cas est fourni par l'évaluation des directions régionales de l'industrie et de la recherche (DRIR) (encadré n° 9) : les monographies menées à cette occasion sur les DRIR Bretagne et Bourgogne ont permis de mieux comprendre les résultats de l'enquête statistique réalisée par ailleurs. Enfin, l'étude de cas est susceptible de mettre en évidence des effets non attendus d'une politique ou des contre-exemples quant à ses effets attendus.

2) La modélisation présente l'avantage de permettre une analyse quantitative fine des effets d'une politique, et notamment de se prêter à la simulation. Ainsi la modélisation élaborée à l'occasion de l'évaluation de l'effet de l'allocation de parent isolé²⁸ sur l'offre de travail a-t-elle permis de simuler les conséquences d'une modification des conditions de versement de cette prestation. A cet égard, les modèles constituent un complément utile aux méthodes typologiques s'appuyant sur des tableaux croisés dans la mesure où celles-ci ne permettent ni de simuler, ni de quantifier des effets, ni de résoudre le problème du lien de causalité.

Le recours à la modélisation se heurte toutefois à divers obstacles :

— il s'agit d'une démarche nécessitant des compétences spécifiques notamment en matière de techniques d'estimations économétriques. Elle peut donc être incompatible avec des contraintes de délais serrés et exiger des moyens (en temps et en spécialisation) qui dépassent souvent les possibilités des équipes d'évaluation mises en place dans les ministères. Sa mise en œuvre nécessite donc le plus souvent de faire appel à des chercheurs ;

— elle suppose que des conditions favorables soient réunies en matière de données : elle est peu envisageable si un dispositif de suivi de la politique n'a pas été mis en place préalablement. En ce qui concerne par exemple l'évaluation de l'API, les documents administratifs se sont révélés difficilement exploitables : ainsi, les dossiers des caisses d'allocations familiales sont progressivement apurés alors qu'il est nécessaire de conserver l'information et de reconstituer « l'histoire » des bénéficiaires de l'API ; d'autre part, l'utilisation des seules sources administratives ne permet pas de prendre en compte les individus qui ne bénéficient pas de la prestation ;

— en outre la modélisation « explicative » a été surtout développée dans les domaines qui permettent de faire appel aux enseignements de la théorie économique (par exemple les politiques d'aides aux entreprises ou aux ménages), et, *a fortiori*, les politiques de régulation macroéconomiques. L'acquis paraît plus faible pour les autres sciences sociales²⁹.

En dépit de ces difficultés, la modélisation est à même de faire progresser l'évaluation sur le plan de la rigueur. Certes une évaluation menée sur cette base

28. J.-C. Ray — op. cit.

29. Cette affirmation doit être tempérée dans certains cas. Il demeure que la modélisation économique, notamment macroéconomique, fait l'objet de travaux depuis maintenant une cinquantaine d'années (voir par exemple l'historique figurant dans les deux premiers chapitres de l'ouvrage de P. Artus, M. Deleau, P. Malgrange, *Modélisation macroéconomique*, 1986, Economica).

L'état des lieux en sciences sociales présenté (en 1975) dans l'ouvrage cité de F.P. Scioli, T. Cook par D.S. Van Meter et H.B. Aster (*Causal perspectives on policy analysis*) est révélateur d'un décalage très important.

L'évaluation des ordonnances de janvier 1982 sur les 39 heures
et la 5^e semaine de congés payés

Cette évaluation, réalisée par des chercheurs du Centre de gestion scientifique de l'École des mines de Paris (*) au deuxième semestre 1982, cherchait à préciser comment les ordonnances abaissant la durée hebdomadaire à 39 heures et généralement la 5^e semaine de congés payés avaient concrètement été mises en œuvre.

L'étude s'est appuyée sur un échantillon non représentatif d'une vingtaine d'entreprises, dont les auteurs avaient déjà analysé, dans le cadre d'une précédente étude, les attitudes potentielles vis-à-vis d'une réduction du temps de travail (**).

La dispersion de la durée du travail (hebdomadaire aussi bien qu'annuelle) s'est accrue à la suite des ordonnances (certaines entreprises ont ainsi réduit fortement la durée hebdomadaire du travail pour créer une 5^e équipe). Les modalités de réduction du temps de travail ont été diverses : en particulier, la différence entre durées de travail affichée et effective (temps de pause, heures supplémentaires) a souvent constitué un enjeu important de négociation.

En ce qui concerne les congés, la généralisation de la 5^e semaine a conduit à un resserrement des situations minimales (elles se situaient entre 4 et 5, 4 semaines avant les ordonnances, entre 5 et 5,5 semaines après).

Face à la réduction de la durée du travail, la logique de productivité a été prédominante au sein de l'échantillon même, si certaines créations explicites d'emplois ont pu être constatées dans certaines entreprises. Ceci indique, selon les auteurs, que « la problématique de la réduction de la durée du travail comme remède au chômage n'est pas spontanément inscrit dans les comportements des acteurs sociaux, qu'il s'agisse des entreprises ou des salariés ». En outre, les ordonnances n'ont pas entraîné de réduction significative d'activité.

Il apparaît enfin qu'au delà de ses conséquences économiques, la réduction de la durée du travail a amené les partenaires sociaux à aborder de front des problèmes d'organisation du travail (pauses, règles non officielles de prise de congés, etc.) qui faisaient auparavant l'objet de compromis informels.

(*) M. Pepin et D. Tonneau : *Réglementation et vie des entreprises : mise en œuvre des ordonnances de janvier 1982 sur les 39 heures et la 5^e semaine de congé*, Communication aux quatrièmes journées d'économie sociale, Paris 27-28 septembre 1984.

(**) M. Pepin et D. Tonneau : *Réduction du temps de travail : une typologie du comportement des entreprises*, Économie et Prévision n° 55 1982/4.

**L'évaluation de l'action régionale
du ministère de l'Industrie et de la Recherche**

Par lettre de mission du 11 août 1983, le ministère de l'Industrie et de la Recherche a demandé au Centre d'évaluation et de prospective (CPE) d'effectuer une enquête relative à l'action régionale du ministère et des établissements sous sa tutelle.

L'objectif premier de cette enquête a été de recueillir et d'analyser de manière assez détaillée l'opinion des acteurs : le personnel des DRIR, les entreprises, les agences et les pouvoirs régionaux. Ensuite de confronter leur vision avec les données internationales concernant le développement régional. Enfin de dégager un petit nombre de concepts structurants.

Pour ce qui est du recueil des données, ont été utilisés des consultations, des enquêtes, et des travaux existants.

Les consultations :

Chaque DRIR a été invité à organiser dans ses services un séminaire de deux jours, avec l'ensemble du personnel et à rédiger un compte rendu de quelques pages à annexer au dossier.

Les établissements suivants : ANVAR, AFME, ADEPA, AFNOR, LNE, CNRS, INPI ont été interrogés par écrit.

Chaque président de Conseil régional a reçu une lettre sous couvert du commissaire de la République, lui demandant les actions en cours et les projets de sa région.

Les partenaires sociaux ont également été saisis par lettre, ainsi que les syndicats représentant le personnel du ministère.

Enfin, tous les directeurs du ministère ont été consultés.

Les enquêtes :

Un questionnaire, comportant 18 questions, a été envoyé en 14 exemplaires à chacune des 153 chambres de commerce, en leur demandant de le transmettre à des PMI représentatives de leur circonscription. 1 279 réponses sont parvenues. Le dépouillement en a été assuré par une société de sondages.

Dans deux régions, deux sociétés d'études ont effectué chacune une opération d'évaluation prospective et participative, comprenant une quarantaine d'entretiens en profondeur, dans et hors de l'administration.

Le Conseil général des mines a effectué, pour l'occasion, une comparaison internationale des missions de contrôle technique.

Encadré n° 9 (suite)

Les travaux existants :

L'Inspection générale a mis ses travaux à la disposition du CPE. Deux demi-journées de travail avec l'ensemble des membres de l'Inspection ont également permis de tirer parti de leurs expériences et de leurs conclusions.

Une enquête sur les unités d'œuvre, lancée par la délégation à l'action régionale, afin d'évaluer la quantité de travail affectée par chaque agent à chaque type d'activité.

Les journées régionales industrie, organisées depuis le début 1983 et les travaux d'élaboration des contrats de plan État-région.

Deux séances de travail ont eu lieu avec les délégués régionaux de la recherche et de la technologie, une avec les directeurs régionaux de l'industrie et de la recherche.

Enfin, le chef du CPE a procédé à une cinquantaine d'entretiens avec des acteurs divers.

peut être, tout autant qu'une autre, sujette à critique. Cette critique devra toutefois porter sur le modèle lui-même, c'est-à-dire sur les mécanismes par lesquels la politique menée est réputée agir sur les décisions des agents et affecter l'environnement économique. Si la nature de ces mécanismes est établie solidement, les effets de la politique estimés par le modèle deviennent relativement incontestables. S'il existe, au contraire, plusieurs interprétations de l'interaction de la politique et de son environnement, plusieurs modèles concurrents sont envisageables : l'estimation économétrique peut permettre, dans certains cas, de les départager.

La situation française est actuellement marquée par un bon développement de la modélisation macroéconomique³⁰, mais un retard important en modélisation microéconomique par rapport aux pays anglo-saxons. Par exemple, les évaluations administratives des aides à l'emploi font très peu référence au modèle microéconomique de la firme alors que les travaux anglo-saxons s'appuient beaucoup plus systématiquement sur ce type de formalisation³¹. En la matière, certains travaux de recherche méritent d'être mentionnés, par exemple les modèles « micro-analytiques » qui s'appuient sur le calcul économique et la stratégie des

30. Ceci explique l'importance des travaux d'évaluation de politiques macroéconomiques avec des modèles. Dans ce cas, les outils sont présents (voir IV-6). On notera que ce déséquilibre entre modélisations macro et micro a des effets négatifs en matière d'analyse de politiques publiques. L'usage sans contrepartie des modèles macroéconomiques sur des périodes longues étend de manière abusive leur période de validité, qui est celle du court terme (2 à 3 ans) et où les mécanismes (essentiellement) keynésiens qu'ils retracent sont dominants. Il en résulte que, dans les simulations macroéconomiques, les investissements les plus coûteux et les moins rentables à moyen terme apparaissent souvent comme les meilleurs, que la politique de l'emploi la plus efficace consiste à recruter des fonctionnaires, etc. Cette remarque, de prudence vis-à-vis d'un usage inconsidéré des modèles macroéconomiques en dehors de leur champ de pertinence, ne diminue en rien leur caractère irremplaçable pour évaluer des politiques et des évolutions à court terme.

31. Voir par exemple la synthèse présentée par J.J. Roger. *Les travaux anglo-saxons relatifs à l'évaluation des actions publiques en matière d'emploi (1975-1978)*. (Journées d'études des 3 et 4 octobre 1985, structures du marché du travail et politiques d'emploi.)

Un exemple d'évaluation faisant appel à la modélisation : l'impact de l'allocation de parent isolé sur l'offre de travail (*)

L'allocation de parent isolé (API) a été créée en 1976 afin de garantir à certaines familles monoparentales un minimum de ressources. Elle peut être versée durant une année après le fait générateur de l'isolement ou jusqu'au troisième anniversaire du dernier enfant. Il s'agit d'une prestation qui ne fait pas l'objet d'un consensus : en effet la durée de versement et son caractère de prestation différentielle sont supposés entraîner une désincitation au travail marchand et donc une accoutumance à l'assistance.

Une évolution de l'API sur l'offre de travail a été réalisée par une équipe de recherche de l'université Nancy II. Elle repose sur l'estimation économétrique de deux modèles complémentaires expliquant la probabilité qu'a une mère isolée de travailler selon qu'elle a le droit de bénéficier de l'API (2^e modèle) ou non (1^{er} modèle). L'avantage de cette approche est qu'elle permet de simuler et de quantifier les effets d'une suppression de l'API ou d'une modification des conditions de son versement : elle constitue donc un complément utile à des méthodes de construction de typologies qui ont été utilisées en parallèle.

Les deux modèles font intervenir comme variables explicatives, à côté du taux de salaire et du revenu non salarial, une variable relative au niveau d'éducation et des variables d'écart correspondant au nombre d'enfants.

Il apparaît qu'en cas de suppression de l'API, 57 % des femmes qui en sont actuellement bénéficiaires travailleraient (essentiellement des femmes enceintes de leur premier enfant ou ayant un seul enfant). Les autres verraient leurs ressources diminuer très inégalement suivant les autres prestations dont elles peuvent bénéficier.

En revanche, la baisse d'un tiers du niveau de ressources garanties par l'API n'amènerait que peu de femmes à travailler.

(*) J.-C. Ray : *Mères isolées, API et travail*. Communication aux quatrièmes journées d'économie sociale, organisées par le Laboratoire d'économie sociale et le Commissariat général du plan.

firmes vis-à-vis des aides à l'emploi et qui permettent d'appréhender les effets de substitution qui en résultent entre les diverses catégories d'emplois³².

La modélisation microéconomique constitue, sans nul doute, un axe prioritaire pour le développement de travaux théoriques et appliqués en France, dans les années à venir, en liaison avec l'évaluation des politiques publiques.

II - 6 Conclusion : la boîte à outils de l'évaluation

Sans prétendre à l'exhaustivité, les techniques d'évaluation peuvent être reclassées en six rubriques :

- l'étude de cas ;
- l'analyse des données (notamment la construction de typologies) ;
- les analyses de séries temporelles (en particulier les méthodes du type Box-Jenkins) ;
- les tests sur petits échantillons³³ ;
- les techniques d'analyse multivariée (tests d'analyse de la variance ou de la covariance) ;
- la modélisation avec estimation économétrique sur données agrégées longitudinales, sur coupes transversales ou sur panels.

Le tableau 1 fournit un croisement entre plans de recherche et techniques d'évaluation, avec le système de notation suivant :

00 signifie qu'une technique est bien adaptée au contexte considéré (plan de recherche, origine des données).

0 signifie qu'il peut être intéressant de mettre en oeuvre la technique en question.

On constate, à la lecture de ce tableau, que l'étude de cas est une technique utilisable pour un large éventail de plans de recherche, même si l'observation *ex post* en l'absence de groupe témoin constitue son domaine de prédilection.

Les analyses de séries temporelles doivent surtout être préconisées dans le contexte de comparaisons avant-après sans groupe de contrôle.

Le domaine potentiel d'application de l'analyse des données et de la modélisation microéconométrique est large lui aussi, à condition toutefois qu'il soit possible de créer de la variance au sein de la population étudiée. En outre, la modélisation

32. Cf. G. Ballot : *Concurrence entre catégories de main-d'œuvre et fonctionnement du marché du travail : l'expérience du modèle ARTEMIS*. Journées d'études organisées par l'Agence nationale pour l'Emploi, le Commissariat général du plan et la Délégation à l'emploi. Paris, 3 et 4 octobre 1985.

33. Il s'agit d'une famille assez hétérogène de techniques, d'usage très fréquent. Elle recouvre des tests plus ou moins structurés, allant de la simple comparaison informelle d'indicateurs jusqu'à des tests non paramétriques aux fondements théoriques bien établis.

Encadré n° 11

L'ÉVALUATION DU PLAN TEXTILE

Description sommaire du plan :

Le plan textile visait à remédier à la situation très préoccupante de ce secteur en 1981. Il avait pour but essentiel de rétablir la situation financière des entreprises, de relancer fortement les investissements, tout en diminuant nettement les pertes d'emplois. L'aide reposait sur des contrats emploi-investissement concernant 4 sous-secteurs : les fils et fibres artificielles et synthétiques (FAS), le textile hors maille, la maille, l'habillement. Ces contrats constituaient un système original puisque sous forme d'aide au fonctionnement (diminution des charges sociales), ils avaient pour but d'aider à la fois l'investissement et l'emploi (le montant de l'aide étant fonction des engagements pris en la matière par l'entreprise).

Les différents plans de recherche utilisés :

L'évaluation du plan textile, réalisé par un groupe de travail interministériel, est caractéristique de l'utilisation conjointe de plusieurs plans de recherche :

- l'analyse avant/après : tentative pour déterminer des ruptures de tendances ;
- des comparaisons de groupes non équivalents issus de l'observation :
 - comparaison avec le secteur du cuir-chaussure ;
 - comparaison avec les secteurs textiles belges et italiens ;
 - comparaison entre entreprises aidées et non aidées.
- une analyse macroéconomique reposant sur l'estimation du coût net du plan et sur des simulations sur modèles macroéconomiques, de ses effets sur l'emploi et la compétitivité dans le secteur textile.

Le tableau ci-dessous résume les effets mesurés à l'aide de ces différents plans de recherche :

Plans de recherche	Effets Mesurés					
	Emploi	Investissement	Prix	Situation financière	Parts de marché	Coût net du plan
Avant/après	X	X		Chiffre d'affaires	X	
Comparaison Secteur cuir-chaussure	X	X		Valeur ajoutée Profit Marge brute d'autofinancement	X	
Comp. internationale	X	X			X	
Comp. aidées/ non aidées	X	X	X		X Structure financière	
Analyse macroéconomique	X				X	X

suppose qu'existe(nt) au préalable une (ou des) théorie(s) susceptible(s) d'expliquer les mécanismes par lesquels transitent les effets de la politique évaluée.

Les tests sur petits échantillons sont particulièrement bien adaptés à de nombreux cas de comparaisons de groupes non équivalents issus de l'observation (comparaisons internationales ou régionales notamment).

Le domaine d'application des tests d'analyse de la variance ou de la covariance est également celui de la comparaison de groupes traités et non traités. Ces tests supposent toutefois un nombre suffisant d'observations, que pourra notamment assurer le recours à l'expérimentation vraie ou à la quasi-expérimentation.

La mise en œuvre de ces différentes techniques dépend en outre de l'origine des données utilisables pour l'évaluation, comme l'indique le tableau 2.

L'étude de cas est essentiellement associée à la constitution de monographies. Les autres techniques peuvent être séparées en deux catégories :

— celles qui peuvent s'appuyer sur des statistiques existantes « agrégées » (données nationales, régionales, sectorielles, etc.) : analyses sur séries temporelles (lorsque l'analyse porte sur l'évolution d'un agrégat affecté par la politique), tests rapides sur petits échantillons (en particulier dans le cas de comparaisons internationales ou régionales), simulation *ex post* des effets de la politique à partir de modèles économétriques estimés sur la base d'agrégats ;

— celles qui nécessitent une collecte de données microéconomiques (et donc en général un dispositif de suivi de la politique) : analyse des données, tests d'analyse de la variance ou de la covariance, modélisation et estimation économétrique sur coupes transversales ou sur panels.

Nous proposons, pour conclure, un arbre de décision indiquant les « meilleurs choix » correspondant à certaines conditions d'environnement ou d'objectifs (cf. tableau 3).

La constitution d'un tel arbre présente toujours une certaine part d'arbitraire : tout dépend en effet du point de vue adopté pour expliquer le choix de l'évaluation (question posée, existence d'un dispositif de suivi préalablement à la mise en œuvre de la politique, type de données disponibles, possibilités de constituer des groupes de contrôle, contraintes de délais, etc.).

L'arbre de décision présenté dans le diagramme ci-dessous n'échappe pas à ce genre de critique ; il privilégie les considérations suivantes :

— le type de question que se pose l'évaluateur (comprendre les mécanismes de fonctionnement de la politique ou mesurer ses effets réels) ;

— la possibilité de constituer des groupes de contrôle ;

— l'existence d'une théorie expliquant les effets de la politique ;

— la possibilité de créer de la variance.

Il ne tient pas compte, en revanche, des contraintes de coût et de délais qui peuvent limiter le choix de l'évaluateur. Les conséquences de telles contraintes seront analysées dans le chapitre suivant, relatif à la mise en œuvre de l'évaluation.

TABEAU 1 : PLANS DE RECHERCHE ET TECHNIQUES D'ÉVALUATION

TECHNIQUES \ PLANS DE RECHERCHE		Pas de groupe de contrôle		Comparaison de groupes non équivalents		Expérimentation avec tirage au sort
		Observation <i>ex post</i>	avant/après	Observation <i>ex post</i>	avant/après	
Etude de cas		00	0	0	0	
Analyse des données		00 (a)	0 (a)	0 (a)	0 (a)	0
Analyse des séries temporelles			00		0	0
Tests sur petits échantillons				00	00	
Techniques d'analyse multivariée				00 (a)	00 (a)	00
Modélisation et estimation économétrique	Sur données agrégées (c)		0 (b)			
	Sur coupes transversales	0 (a) (b)		00 (a) (b)		00
	Sur panels		0 (a) (b)		00 (a) (b)	00 (a) (b)

(a) A condition qu'il soit possible de créer de la variance.

(b) Suppose un schéma théorique préalable.

(c) On peut classer dans cette rubrique la simulation des effets d'une politique à partir de modèles macroéconométriques.

TABLEAU 2 : TECHNIQUES D'ÉVALUATION ET ORIGINE DES DONNÉES

ORIGINE DES DONNEES TECHNIQUES	Monographies	Statistiques globales existantes (pas ou peu de variance)	Données microéconomiques	
			Issues d'enquêtes statistiques	Issues d'expérimentations ou de quasi expérimentations
Etude de cas	00			
Analyse des données			0	0
Analyse de séries temporelles		00 (a)	0 (b)	0 (b)
Tests sur petits échantillons		0 (c) (d)		
Techniques d'analyse multivariée			0 (c)	00
Modélisation et estimation économétrique		0 (a)	00	00

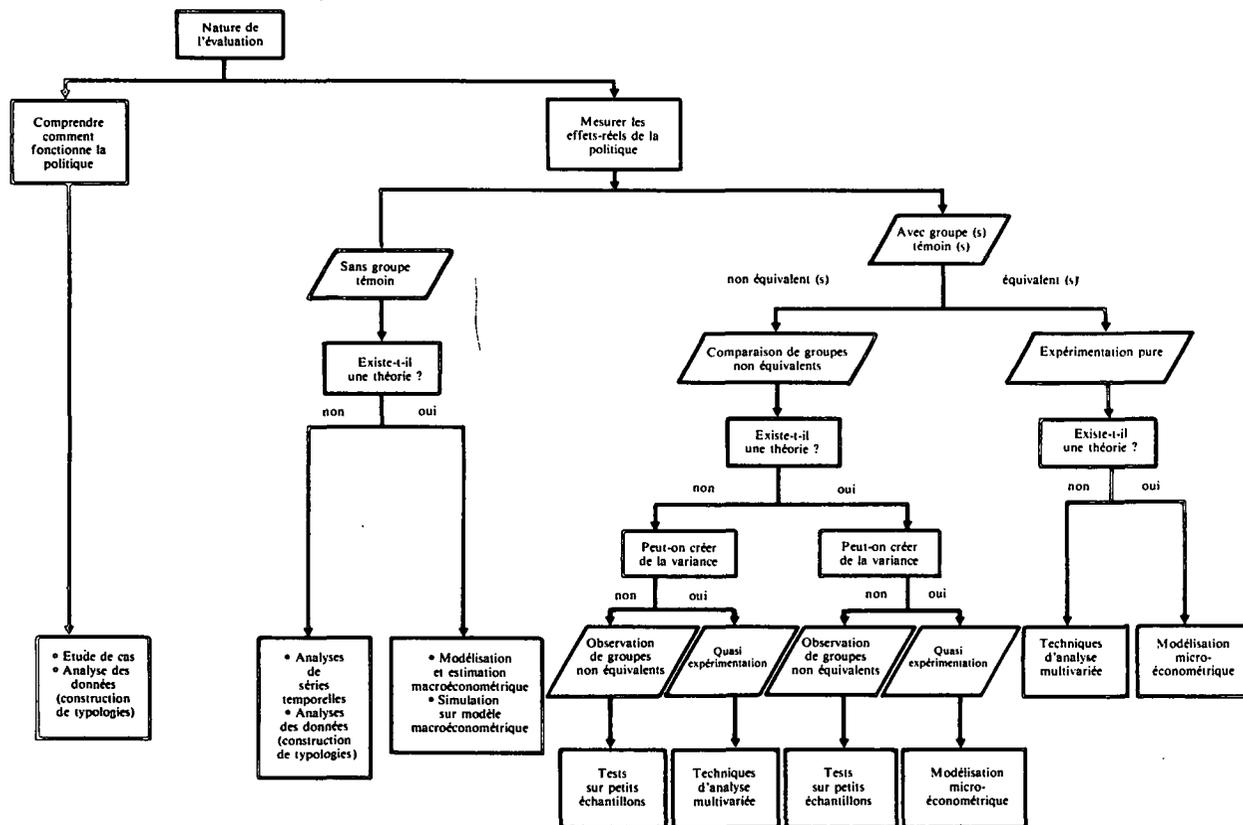
(a) Si l'on dispose de données longitudinales.

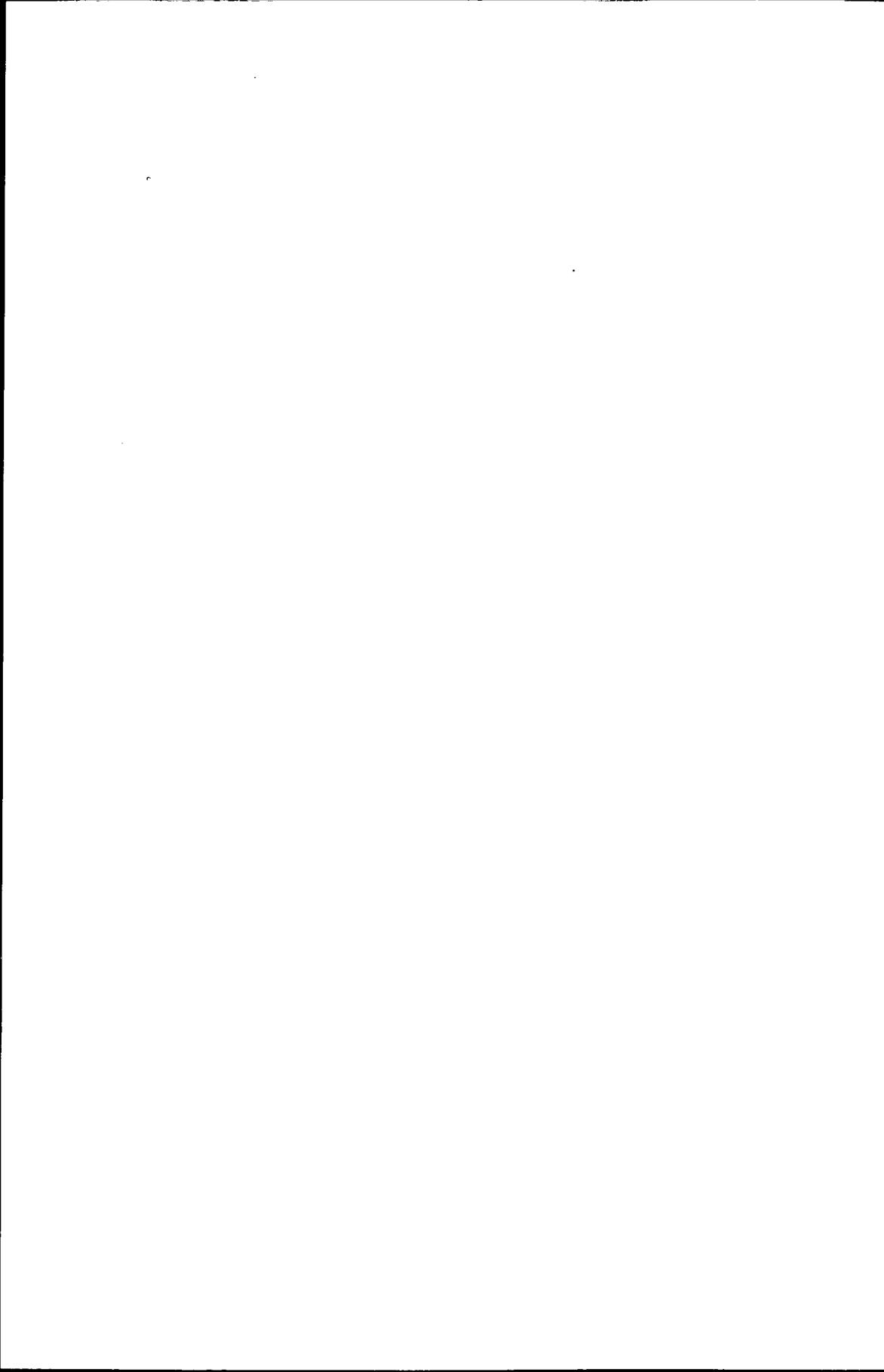
(b) S'il s'agit de données sur panels.

(c) Si l'on peut distinguer entre groupe(s) cible(s) et groupe(s) témoin(s).

(d) Comparaisons internationales et régionales notamment.

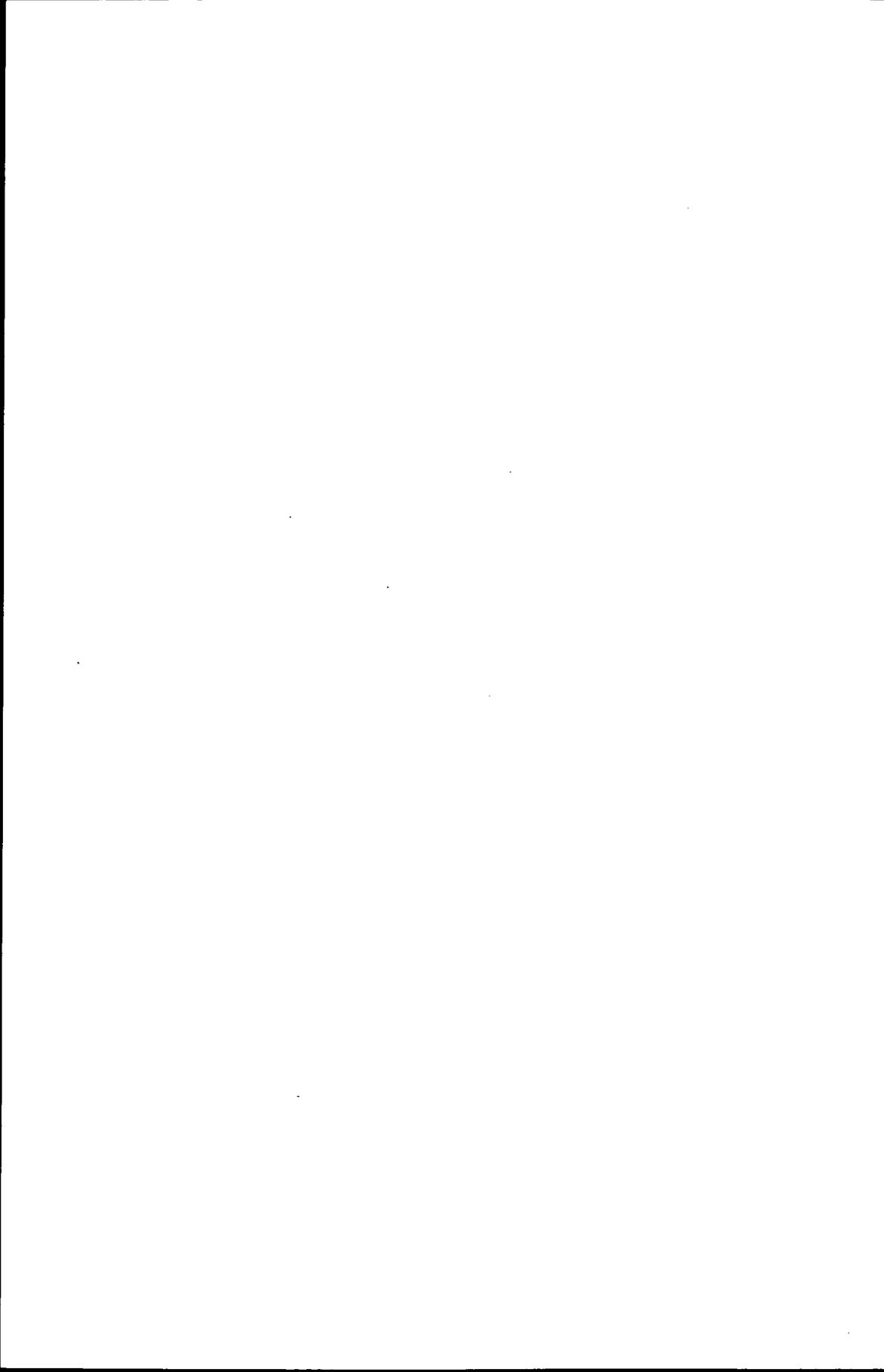
**TABEAU 3 : PLANS DE RECHERCHE ET TECHNIQUES DE L'ÉVALUATION
UN ARBRE DE DÉCISION**





III

La mise en œuvre de l'évaluation



III

La mise en œuvre de l'évaluation

III - 1 Introduction

Les développements précédents ont concerné la **boîte à outils** de l'évaluation : plans de recherche alternatifs, techniques utilisables pour le traitement et l'interprétation des données collectées. Nous nous intéressons maintenant au **mode d'emploi** de cette boîte à outils. Autrement dit, il s'agit d'élaborer quelques propositions sur l'utilisation, **en situation**, de ces plans de recherche et de ces techniques : dans telle configuration d'évaluation, existe-t-il un meilleur plan de recherche ? Y-a-t-il au contraire certains arbitrages possibles etc. ? Telles sont les questions auxquelles un bon mode d'emploi devrait répondre.

Là encore il n'est pas envisageable, ni souhaitable, de fournir un guide *presse-bouton* qui aplanirait toutes les difficultés et résoudrait une fois pour toutes les problèmes que pose à l'évaluateur le choix de sa démarche. Chaque situation concrète appelle des ajustements spécifiques et même, souvent, l'utilisation combinée de plusieurs approches. On peut toutefois dégager quelques grands principes¹.

La mise en œuvre de l'évaluation est conditionnée par **deux facteurs principaux** :

— **la nature des questions posées** : s'interroge-t-on seulement sur la manière dont la politique a été appliquée (question descriptive aboutissant à une évaluation de mise en œuvre) ? Cherche-t-on en outre à apprécier ses effets spécifiques (question en forme de causalité demandant une évaluation de résultat) ? Veut-on, ou non, étendre les conclusions (problème de la validité externe) ?

— **les contraintes et limites** de diverses natures, qui sont des données pour l'évaluation, en matière notamment :

- de collecte de l'information ;
- de coût ;
- de délais ;
- d'insertion administrative.

Compte tenu des questions posées et des contraintes auxquelles il est confronté, l'évaluateur doit réaliser certains arbitrages « stratégiques », avant de procéder à l'évaluation proprement dite :

1. La démarche qui suit recoupe certains développements contenus dans l'ouvrage de méthode du GAO, *Designing evaluations* (op. cit.).

— arbitrages en matière d'objectifs de l'évaluation : à quelles questions est-il susceptible de répondre, et avec quel degré de validité ?

— arbitrages relatifs au sujet de l'évaluation, c'est-à-dire le champ examiné et le niveau auquel il se situe : faut-il chercher à évaluer des actions ponctuelles ou au contraire des politiques très globales ? Peut-on se placer à un niveau intermédiaire d'objectifs de l'action publique ?

Une fois ces arbitrages de nature stratégique effectués (paragraphe III-2), il reste à choisir le(s) plan(s) de recherche et la (les) technique(s) d'exploitation des données (paragraphe III-3). Au-delà de la phase d'évaluation proprement dite il convient encore de s'interroger sur « l'après-évaluation » (paragraphe III-4) : les conclusions obtenues permettent-elles de conforter, de réorienter, ou de remettre en cause certaines actions de la puissance publique ? Est-il souhaitable de poursuivre l'évaluation de la politique considérée en précisant le sujet et en se dotant de davantage de temps et de moyens ? Enfin, quelques règles simples de déontologie générale seront proposées (paragraphe III-5).

Les développements qui suivent fournissent ainsi des recommandations générales pour la conduite d'une évaluation. Comme il a été dit, ils ne peuvent viser à l'exhaustivité. En particulier, il serait artificiel d'émettre des conclusions figées sur la prise en compte de l'environnement « humain » de l'évaluation. On ne peut se borner, en la matière, qu'à une recommandation d'attention. Il est clair que l'évaluateur occupe souvent une position très particulière au sein d'un réseau d'acteurs : il est (ou il doit être — voir III-5) distinct de ceux qui décident la politique ou l'exécutent. Il peut alors être conduit à rechercher de l'information pertinente auprès de ces acteurs. Plus généralement, l'évaluation peut impliquer une multiplicité d'agents concurrents en termes de pouvoir ou de problématique, à l'intérieur ou à l'extérieur de l'administration. Deux écueils se présentent pour l'évaluateur. Le premier consiste évidemment à ignorer superbement cet environnement et à adopter l'attitude de celui qui sait. On risque, ce faisant, à la fois de rencontrer des difficultés importantes pour conduire l'évaluation et d'aboutir à un rejet de ses résultats. Le second est symétrique et consiste en l'abandon de toute position externe de l'évaluateur et, en particulier, de sa responsabilité finale, qu'il détient et ne peut céder : **émettre un jugement sur les résultats**. Cette démission conduit directement à une digestion de l'évaluateur par le système et l'acte d'évaluation dégénère en animation de groupe ou en vote de motion de synthèse.

Au total, et parce qu'il faudra bien dans de nombreux cas solliciter les acteurs pour analyser et comprendre la politique et répondre à ceux qui chercheront à exercer une pression sur l'évaluation, une voie moyenne peut être proposée. Cette voie suppose la fixation d'un **contrat** préalable et l'adoption d'un protocole adapté. Le contrat est d'autant plus facile à conclure que la procédure d'évaluation est définie à l'amont, dès la formulation de la politique à évaluer. Une telle démarche contractuelle peut être plus délicate à suivre lorsqu'il s'agit d'étudier des politiques bien « enracinées », génératrices de rentes pour certains acteurs. Dans ce cas, la recherche d'un consensus ne saurait servir d'excuse à l'immobilisme et les risques d'un conflit doivent être affrontés. De même, le protocole adopté, notamment à propos de la mise à contribution des acteurs pour produire et réunir des informations sur la politique, doit faire la part de la complexité du sujet traité, de

ses composantes subjectives. Si une attitude d'écoute est à recommander, il convient de ne pas perdre de vue l'objectif. Celui-ci n'est pas de faire comprendre aux acteurs que leurs préoccupations sont prises en compte, il est de produire un jugement sur la politique suivie.

III - 2 La stratégie de l'évaluation

En amont du choix proprement technique d'un plan de recherche particulier, la mise en œuvre de l'évaluation requiert l'examen de **certaines facteurs généraux d'environnement** et demande que soient fixées certaines **orientations stratégiques**. On traitera successivement des points suivants :

- les objectifs de l'évaluation et les contraintes d'environnement ;
- le niveau d'évaluation de la politique ;
- le choix des experts.

1) **Les questions auxquelles l'évaluation des politiques publiques peut être amenée à répondre sont diverses.** En allant du plus simple vers le plus complexe, on peut distinguer :

— des questions de nature **descriptive** :

- comment la politique a-t-elle été mise en œuvre ? et l'a-t-elle été ?
- sur quels instruments était-elle fondée ?
- quels étaient ses objectifs ?
- ces objectifs ont-ils été atteints ?

— des questions impliquant une analyse des **causalités** :

- si les objectifs de la politique ont été atteints, en quoi l'action publique en est-elle responsable ?
- en quoi les différents ingrédients de la politique publique ont-ils contribué à la réalisation de ces objectifs ?
- une autre politique aurait-elle été plus efficace ?
- peut-on généraliser les résultats de l'évaluation à une politique de même nature mais visant d'autres cibles (autres secteurs, autres groupes d'agents) ou se situant dans un autre contexte (une autre région par exemple) ?

— des questions aboutissant à un **jugement de valeur** :

- la politique a-t-elle eu des effets positifs ou négatifs ?
- que peut-on dire *ex post* de la pertinence de ses objectifs ?
- quels autres objectifs devraient éventuellement leur être substitués ?

Encadré n° 1

La mise en œuvre de l'évaluation par le General Accounting Office (*)

Le General Accounting Office (GAO) a été créé par le Budget and Accounting Act de 1921 qui en fait la contrepartie législative du Bureau du Budget (devenu Office of Management and Budget en 1970).

Le GAO est dirigé par un Comptroller General et un Deputy Comptroller General nommés pour quinze ans par le président des États-Unis avec l'accord du Sénat. Le Comptroller General n'est subordonné ni au Président ni au Congrès, et bénéficie du statut d'« Officer of the United States ».

Le principal domaine de compétence du GAO s'étend à toute question relative à l'utilisation des fonds fédéraux. Le GAO a un droit de regard sur tout document administratif et a toute latitude pour retenir le niveau qui lui paraît le plus adapté (simple contrôle financier, ou analyse d'efficacité).

Les évaluations réalisées par le GAO s'appuient sur des relations négociées avec les administrations, pour l'obtention des informations nécessaires et la mise en œuvre des réformes préconisées. La présence physique des évaluateurs du GAO au sein des administrations facilitent ces relations.

Le GAO est en fait très indépendant. Il accorde certes une certaine priorité aux demandes d'évaluation présentées par le Congrès, mais il a également ses propres priorités. Dans l'ensemble le GAO est très bien perçu par le Congrès.

Les autres compétences dont dispose le GAO ne sont pas sans lien avec l'évaluation : amélioration des systèmes comptables des administrations, interprétation des textes en matière de finances publiques ; conseil auprès du Congrès sur les questions législatives, réglementaires et juridiques ; compétence juridictionnelle contentieuse pour connaître notamment les litiges financiers portés par ou contre le gouvernement des États-Unis.

La « comprehensive evaluation » et sa mise en œuvre :

La « comprehensive evaluation » regroupe les différents types d'évaluations mises en œuvre par le GAO :

- l'audit des opérations financières et des systèmes de comptabilité ;
- l'audit d'« efficacité et d'économie » qui permet de savoir comment sont gérés les programmes ;
- l'évaluation portant sur l'atteinte des objectifs des programmes ;
- l'évaluation dégageant des approches alternatives à partir d'analyses coûts-avantages ;
- des « études spéciales » : recherches méthodologiques, promotion de nouvelles techniques de gestion.

(*) Voir la communication au groupe de Luc Rouban, « L'évaluation des politiques publiques aux États-Unis ».

Encadré n° 1 (suite)

La « comprehensive evaluation » comprend à la fois des évaluations visant à apprécier si une politique a été correctement mise en œuvre (process evaluations) et des évaluations ayant pour objet de mesurer l'effet d'une politique et son efficacité (impact evaluations). Le GAO a cherché à mieux articuler ces deux types d'évaluations bien qu'il ait tendance en raison du contexte institutionnel dans lequel il fonctionne (évaluations faites à la demande du Congrès, supervision de la gestion des administrations) à mettre davantage l'accent sur les « process evaluations ».

L'ampleur d'une évaluation réalisée par le GAO dépend :

- de l'intérêt qu'il trouve à étudier la question posée ;
- de l'existence préalable d'informations (notamment de rapports précédents d'évaluation) ;
- de l'éventualité d'un risque de conflit avec le Congrès ou l'Administration pour des raisons de nature politique.

Selon l'objet de l'évaluation, le GAO peut avoir recours à toutes les techniques possibles. Toutefois, le fait qu'il intervienne souvent « au second degré », à partir d'évaluations réalisées par les administrations responsables et dont il peut vérifier le protocole, explique qu'il ne fasse que rarement appel aux méthodes de quasi-expérimentation ou d'expérimentation pure, ou encore à l'analyse sociologique.

Les rapports d'évaluations réalisées par le GAO ne sont pas tous rendus publics : certaines études restent confidentielles.

La mise en œuvre de la « comprehensive evaluation » nécessite des moyens importants : le GAO dispose de 5 000 agents répartis entre l'état-major de Washington, 80 bureaux d'audits installés dans certaines administrations, 15 bureaux régionaux et 3 bureaux internationaux.

La sélection des objectifs de l'évaluation est faite à partir de vingt-neuf aires de recherche (« issue areas »). Chaque thème est affecté à un département du GAO qui établit un plan de travail sur une base de deux à quatre ans.

Le personnel du GAO est opérationnel à 80 %. Il comporte des spécialistes de différents domaines (comptabilité, droit, économie, gestion, informatique, etc.).

La plupart des départements du GAO établissent en outre des listes d'experts extérieurs sur des sujets spécialisés. Deux groupes d'experts, l'un composé de membres du secteur privé et de l'administration, l'autre d'universitaires et d'experts en sciences sociales, ont en outre un rôle consultatif.

Le prescripteur de l'évaluation attend bien évidemment de celle-ci qu'elle lui fournisse le maximum de réponses aux questions qu'il se pose sur l'efficacité de la politique étudiée. Il attend en outre que ces réponses lui soient apportées à temps pour que l'action menée puisse être, si nécessaire, infléchie ou révisée. Il est enfin tenu de veiller à ce que le coût de l'évaluation reste dans les limites qu'il juge raisonnable.

Le problème de l'évaluateur est donc « d'optimiser » la finesse et la robustesse de ses conclusions sans pour autant dépasser les limites qui lui sont fixées, notamment en matière de temps et de budget. En règle générale, aller au-delà du constat simplement descriptif et approfondir les analyses de causalité exige de recourir à des techniques sophistiquées s'appuyant sur des dispositifs relativement lourds et coûteux : le souci de validité entre donc naturellement en conflit avec les contraintes qui s'imposent à l'évaluateur.

Il serait injustifié de tirer de cette constatation des conclusions exagérément pessimistes. Certes, dans bien des cas, le poids des contraintes obligera l'évaluateur à se fixer des objectifs plus modestes qu'il ne l'aurait souhaité, soit en restreignant son champ d'investigation, soit en recherchant seulement des présomptions et non pas des preuves réfutables quant aux effets de la politique. Pour des raisons déontologiques, une telle acceptation de résultats incomplètement fiables serait inadmissible dans certains domaines d'action. En matière de thérapeutique notamment, il est inconcevable de décider la diffusion d'un médicament ou d'un traitement sans que les effets et l'efficacité en aient été décrits complètement dans le détail. Dans d'autres domaines, en revanche, celui des politiques sociales en particulier, il peut être tout à fait judicieux de se contenter d'une rigueur moins absolue² : ainsi, pour le décideur, l'évaluation de politique qui établit de simples présomptions est déjà significative, surtout si elle permet d'infléchir l'action en temps utile.

En perdant en validité³ pour gagner en rapidité, l'évaluation permet dans certains cas d'améliorer l'efficacité de l'action publique. A la limite, une évaluation menée avec des techniques sophistiquées, et qui aboutit alors que la politique considérée s'est déroulée complètement, peut n'avoir aucun intérêt autre que historique. A l'opposé, une évaluation de mise en œuvre menée assez près de l'action elle-même et avec des techniques rustiques peut permettre d'arrêter ou d'infléchir une politique ayant des effets pervers inattendus. Toutefois, cette dernière façon de faire n'est pas compatible avec la prise en compte d'effets indirects sur longue période.

Les observations précédentes ne signifient évidemment pas que l'exigence de rigueur doive être volontairement négligée : il convient au contraire que l'évaluation soit aussi rigoureuse que possible, dans le respect des contraintes dans lesquelles elle s'insère. Elles suggèrent plutôt une démarche de réponse graduée :

2. La démonstration totalement rigoureuse étant par ailleurs souvent impossible dans un milieu où les mécanismes en cause sont de nature aléatoire et évolutive.

3. Il est probable que les réflexes de rejet vis-à-vis d'évaluations perçues comme « sales » sont assez forts dans les milieux scientifiques français, moins familiers des démarches pragmatiques que leur homologues américains. D'où les reproches d'« académisme » formulés parfois par les décideurs administratifs à l'encontre d'évaluations réalisées par des centres d'études et de recherche.

— dans un premier temps, il convient de tirer tous les renseignements possibles des techniques les plus faciles à mettre en œuvre. La **diversification des plans de recherche**, non affectés par le même biais, est susceptible de pallier le faible degré de validité de chacun d'entre eux, pris individuellement, et peut permettre, dans certains cas, d'aboutir à des présomptions fortes quant aux effets de la politique ; une telle démarche a été mise en œuvre lors de l'évaluation du plan textile ;

— si ces présomptions paraissent trop sujettes à caution, une évaluation complémentaire menée avec des moyens plus importants (dispositif d'observation, approche expérimentale, utilisation de techniques statistiques complexes, etc.) peut être alors déclenchée. Cette deuxième phase exige des délais plus longs et peut nécessiter une certaine implication des services opérationnels concernés, ainsi qu'un appel plus massif à des compétences extérieures à l'administration (des laboratoires universitaires à caractère « permanent » dont on cherche à déterminer les ingrédients les plus appropriés (aides à des secteurs en difficulté, aides à l'emploi, etc.) à partir d'une analyse approfondie du comportement des acteurs.

Enfin, l'importance de la nature des informations disponibles va conditionner étroitement la mise en œuvre de l'évaluation :

— **le contexte le plus favorable correspond à l'existence d'un dispositif de suivi de la politique.** Bien que ce cas ne soit pas le plus fréquent dans notre pays, il n'est pas exceptionnel (insertion des jeunes de 16 à 18 ans, sécurité routière, périnatalité, TGV Sud-Est, travaux d'utilité collective, préretraites) et il semble se développer (souci de mettre au point un dispositif permanent en matière d'aides à l'emploi). Aux États-Unis, la mise en place d'un dispositif de suivi intervient presque systématiquement lors du lancement d'un nouveau programme social.

L'existence d'un tel dispositif permet de mener des analyses approfondies (recours aux méthodes statistiques) et présente l'avantage de fournir des observations d'unités individuelles diversifiées.

Certains de ces dispositifs autorisent en outre des comparaisons de groupes non équivalents (quasi-expérimentation menée dans le cadre des stages de formation des jeunes de 16 à 18 ans par exemple).

— **en l'absence de dispositif de suivi, il convient de ne pas négliger les informations provenant des systèmes d'information existants, en particulier les fichiers de l'administration.** Leur utilisation pose toutefois une série de problèmes : secret statistique, destruction de l'information après clôture de l'exercice (difficulté rencontrée notamment dans le cadre de l'évaluation de l'API), manque de normalisation, etc. S'ils donnent une bonne image de la population atteinte, ces fichiers ne permettent pas, en revanche, de constituer des échantillons représentatifs de l'ensemble de la population visée.

Par ailleurs, l'existence d'organismes relais qui ont en charge tout ou partie de l'action menée (le Crédit Agricole pour les prêts bonifiés à l'agriculture par exemple) introduit un degré de complexité supplémentaire et peut rendre plus difficile la mobilisation de l'information nécessaire à l'évaluation.

Le système statistique (INSEE, services statistiques des ministères) fournit des informations stables et qui peuvent être utiles à l'évaluation de politiques publiques même si elles ne sont pas collectées à cette fin. Les grands fichiers (répertoires d'entreprises, recensement) permettent en particulier de tirer des échantillons

représentatifs nécessaires à la mise en place d'un dispositif d'évaluation (l'enquête auprès des entreprises sur les aides à l'emploi menée dans le cadre de l'étude prioritaire RCB sur ce thème s'est appuyée par exemple sur un échantillon d'entreprises demandé à l'INSEE).

2) Le manque de transparence des objectifs de certaines politiques publiques a été évoqué à plusieurs reprises au sein du groupe pour illustrer les obstacles à l'évaluation. Ce manque de transparence peut avoir plusieurs origines :

— l'imprécision du contenu de la politique

La politique de santé, par exemple, ne fait pas directement référence, dans ses objectifs, à la notion de santé. Elle est définie en termes d'organisation et de gestion de l'offre de soins et de prévention. Or, mesurer les effets d'une politique d'offre de soins ne permet pas toujours d'en déduire son impact sur la santé des individus. C'est négliger les mécanismes complexes afférents à l'ajustement de l'offre à la demande. Ainsi, en l'absence d'un équipement lourd donné, un hôpital dirige ses malades vers un centre spécialisé non saturé et mieux qualifié. Lorsqu'il aura installé ce matériel, il permettra de mieux localiser les traitements mais ceci n'aura qu'un effet réduit sur les malades eux-mêmes.

— l'existence d'objectifs non affichés

Ce cas de figure est assez fréquent :

- les politiques agricoles ne comportent pas d'objectifs explicites sur le nombre des agriculteurs ; et pourtant le maintien d'un milieu rural constitue une préoccupation réelle et permanente des décideurs ;
- de même, les aides à l'emploi qui accompagnent les restructurations industrielles ont notamment pour vocation de limiter les conflits sociaux pouvant survenir à cette occasion ; cette ambition reste largement implicite ;
- quand aux objectifs de certaines politiques d'emploi, ils paraissent pour le moins ambigus. Suivant quel critère apprécier les stages destinés aux jeunes : l'insertion dans la vie active, l'amélioration de la qualification, la lutte contre la marginalisation irréversible, ou encore l'impact sur les statistiques du chômage ? Les résultats sont souvent limités vis-à-vis des deux premiers objectifs : les taux de placement après le stage restent faibles, les stages de qualification sont minoritaires. Les deux autres objectifs n'ont pas été affichés explicitement. L'un se situant dans le moyen terme échappe à des évaluations courtes, tandis que l'autre a probablement une priorité égale aux yeux des décideurs politiques, ce qui explique le recours systématique à de telles politiques dans la plupart des pays.

— la coexistence d'objectifs antinomiques, au moins à court terme

L'existence d'objectifs contradictoires est également assez fréquente. Elle se manifeste par exemple dans le cadre de politiques d'aides sectorielles à l'industrie qui cherchent d'une part à freiner l'hémorragie d'emplois dans un secteur et, d'autre part, à améliorer sa compétitivité, ce qui peut supposer une réduction d'effectifs. Cette ambivalence existait dans le cas du plan textile. Les politiques visant à améliorer la productivité des exploitations agricoles françaises vis-à-vis de l'étranger tout en limitant la diminution du nombre d'agriculteurs fournissent une autre illustration de ce phénomène.

Analyser les objectifs d'une politique est donc un exercice délicat ; l'évaluateur peut être conduit à cette occasion à reconstituer *ex post* et au prix de beaucoup d'efforts un regroupement et une cohérence qui n'existaient pas nécessairement auparavant. Bien que chaque expert ait souvent tendance à considérer ces difficultés comme propres au secteur dont il est chargé — pour le spécialiste de l'agriculture, l'action publique apparaît plus facile à cerner en matière d'emploi ou de santé... et réciproquement —, les travaux du groupe ont montré qu'il s'agit d'un problème général qui appelle la mise en œuvre d'instruments largement communs. L'analyse sous forme de structure d'objectifs — programmes (S.O. — P.)⁴, peut ainsi mettre en évidence les différents niveaux d'objectifs de la politique étudiée. Elle constitue sans doute un point de départ pour « débroussailler » un sujet complexe dont la cohérence interne n'apparaît pas au premier coup d'œil. Ce type de démarche a par exemple été retenu dans le cadre de l'évaluation des aides à l'élevage bovin ; il a permis d'établir l'existence de trois niveaux différents, deux d'objectifs et un d'action (cf. encadré 1 — partie I) :

- **les objectifs généraux** (finalités de la politique) ne sont pas spécifiques à l'élevage bovin : participation à la croissance, redressement du commerce extérieur, sécurité d'approvisionnement, augmentation du revenu des agriculteurs, équilibre du milieu rural ;
- **les objectifs opérationnels** portent plus spécifiquement sur la filière bovine : orientation des productions, organisation des producteurs, amélioration de la productivité du cheptel et de la qualité des produits, etc. ;
- à ces objectifs opérationnels correspondent **des actions** et des moyens qui ont été regroupés en programmes : amélioration génétique, prophylaxie, interventions sur les marchés, etc.

Cette décomposition — moyens → programmes → finalités — conduit à privilégier le niveau des objectifs opérationnels, ce qui permet d'éviter à la fois le caractère flou des évaluations de politiques globales et l'absence de vision cohérente de l'action publique qui risque de caractériser une évaluation menée à un niveau plus fin (celui des programmes élémentaires constituant la politique). Si le pouvoir explicatif de l'évaluation s'en trouve amélioré, il se peut toutefois que ce soit au prix d'un affaiblissement de sa problématique : l'évaluateur s'efforce de voir si les objectifs opérationnels ont été atteints et si la politique suivie en est responsable, il renonce par contre le plus souvent à juger *ex post* de la pertinence de ces objectifs. Prenons l'exemple du plan textile : il est certes plus facile d'appréhender ses effets en termes d'emplois dans le secteur textile (objectif opérationnel) que de mesurer son impact sur l'emploi global (finalité), notamment parce qu'un emploi créé a un endroit ou dans un secteur est peut-être un emploi délaissé ailleurs. Toutefois en se limitant aux effets internes au secteur textile, l'évaluation permet déjà de juger de l'adéquation du plan aux objectifs opérationnels visés (emplois et compétitivité du secteur) ; elle ne permet en aucun cas de dire si l'État a eu raison, du point de vue de ses objectifs globaux en matière d'emploi et de compétitivité de l'économie, de soutenir le textile plutôt qu'un autre secteur.

4. Pour reprendre une dénomination employée dans les premiers temps de l'opération RCB.

On peut également observer une certaine confusion entre les objectifs internes ou externes en particulier dans le domaine des politiques sociales. Ainsi par exemple, l'évaluation de tel programme éducatif sera très différente si l'on cherche à évaluer la façon dont les élèves assimilent le programme (point de vue interne) ou si on évalue l'utilité des capacités acquises pour la vie professionnelle (point de vue externe). De la même manière, il est différent d'évaluer les TUC par rapport à la situation d'emploi ou par rapport aux conditions de vie qu'ils permettent aux intéressés. Dans les deux cas les politiques visent simultanément les deux catégories d'objectifs mais l'évaluation peut légitimement privilégier l'un ou l'autre objectif. Le problème est de procéder à l'analyse préalable des objectifs et d'afficher les choix sur lesquels repose l'évaluation.

Ce type de difficultés risque d'apparaître chaque fois que la contribution des objectifs opérationnels aux finalités de l'action de l'État ne semble pas aller de soi.

3) **La réalisation de l'évaluation nécessite de recourir à des experts.** Le choix de ceux-ci doit respecter un certain nombre de principes :

— ces experts doivent être externes par rapport à la politique à évaluer. Ce principe de séparation entre l'évaluation et l'exécution (cf. partie III-5) est valable, que l'évaluation soit elle-même administrative ou non. Ainsi le service d'évaluation du ministère de l'Éducation nationale⁵ est-il directement rattaché au ministre, et donc indépendant des directions opérationnelles responsables de l'élaboration, des programmes d'enseignements. Toutefois, la réalisation de l'évaluation se fait avec la collaboration des enseignants et sur la base de critères issus des directives de l'Inspection générale.

— le groupe d'experts réunit des spécialistes de la politique à évaluer et des spécialistes des méthodes d'évaluation. Il convient d'une part d'éviter le danger d'imprécision de l'évaluation liée à une familiarité insuffisante avec les plans de recherche et les techniques utilisables et, d'autre part, de veiller à ce que le groupe puisse disposer de toutes les informations nécessaires sur la politique à évaluer. La présence, au sein du groupe, d'experts ayant une compétence technique pointue tant dans les différentes disciplines concernées (économie, droit, psychologie, sociologie,...) que sur le sujet traité, peut-être indispensable. Il en est ainsi, notamment, lorsque l'évaluation fait appel à la modélisation ou à des méthodes statistiques complexes. Il est en outre très souhaitable d'associer ces experts à la mise en place du dispositif d'évaluation, de manière à assurer une cohérence entre ce dispositif et les conditions de mise en œuvre de l'évaluation (type d'informations, nature des données).

— la désignation du groupe d'experts appartient à l'autorité qui a le pouvoir de déclencher l'évaluation ; le mandat confié au responsable du groupe doit définir implicitement le contenu minimal de l'évaluation sans pour autant limiter de façon excessive la réflexion, celle-ci doit pouvoir se développer sur tous les sujets et les problèmes et utiliser tous les instruments susceptibles d'affermir l'évaluation entreprise.

5. Le SPRESE : voir encadré n° 3, partie II.

Encadré n° 2

Mise en œuvre de l'évaluation et choix des experts : l'exemple du comité national d'évaluation des universités

Le Comité national d'évaluation des universités est un organe spécifique chargé d'évaluer les établissements d'enseignement supérieur de l'Éducation nationale. Cette évaluation, qui constitue la contrepartie de l'autonomie accordée aux établissements d'enseignement supérieur, est réalisée établissement par établissement et par thèmes transversaux.

Le Comité se compose de 15 personnalités nommées pour quatre ans et d'un secrétariat général.

Les établissements faisant l'objet d'une évaluation seront découpés en unités. Pour chaque unité, le Comité enverra un groupe de deux ou trois experts, choisis par lui-même parmi des enseignants, des chercheurs, des cadres d'entreprises et des personnalités étrangères. Ces experts devront posséder une certaine légitimité : ils doivent être jugés par les responsables des unités étudiées comme ayant toutes les capacités requises par l'évaluation.

Le Comité cherchera à réunir toute l'information possédée par l'administration centrale afin de fournir aux experts une première synthèse qui pourra éclairer sa mission.

III - 3 Le choix d'un plan de recherche

Une fois l'environnement « stratégique » de l'évaluation exploré et défini reste une question de logistique importante : **le choix d'un (ou de) plan(s) de recherche et des techniques afférentes**. Ce choix va dépendre de divers facteurs, dont les suivants :

- la nature descriptive ou causale de la question posée. S'agit-il d'une évaluation de mise en œuvre ou d'une évaluation de résultat ?
- le souci de généraliser les conclusions de l'évaluation ;
- les contraintes de délais et de budget de l'évaluation ;
- la possibilité de multiplier les observations en coupe transversale sur des sujets assez différents les uns des autres ;
- l'existence d'observations de même nature avant et après mise en œuvre de la politique ;

— la possibilité de trouver ou de constituer des groupes de comparaison et éventuellement de recourir à l'expérimentation avec tirage au sort ;

— l'existence d'une réflexion théorique permettant l'élaboration de modèles.

Cette liste de facteurs fait apparaître une multiplicité des cas de figure possibles et montre qu'il serait illusoire en l'état actuel des choses de chercher à associer de manière systématique un plan de recherche à une évaluation d'un type donné. Un certain nombre de propositions générales peuvent toutefois être suggérées.

○ En dehors des contraintes de coût, de délais et de mise en œuvre, le choix d'un plan de recherche dépend essentiellement des exigences de l'évaluation :

— lorsque la question posée est simplement descriptive, une bonne validité de la réponse pourra se satisfaire de dispositifs simples ; les plans de recherche sans comparaison avec des groupes témoins sont alors assez adaptés ;

— si l'on cherche en plus à savoir dans quelle mesure la politique est à l'origine des effets observés, le souci de validité conduit à comparer des groupes affectés et des groupes non affectés par la politique. Si ceci s'avère impossible, des analyses avant-après peuvent toutefois apporter certains enseignements d'autant plus précieux qu'ils résultent de l'observation de panels ;

— enfin, le souci de généraliser les conclusions de l'évaluation à d'autres cibles (autres secteurs, autres groupes d'agents) ou à d'autres contextes (autres pays, autres régions, etc.) demande que soit déterminé l'ensemble des facteurs qui peuvent interférer avec la politique et que les effets en soient mesurés. Ceci passe de préférence par des comparaisons entre groupes et peut donner lieu à une modélisation plus ou moins poussée. Pour certaines politiques caractérisées par l'impossibilité ou la grande difficulté des comparaisons entre groupes, des approches plus simples pourront tout de même mettre en évidence des facteurs particulièrement favorables (ou défavorables) à la réussite de la politique (l'exemple de la prophylaxie de la brucellose, évoqué précédemment, est significatif de ce point de vue).

● La deuxième conclusion est que des contraintes très fortes de coût et de délais conduisent à privilégier des approches rapides et peu approfondies et rendent difficile la mise en place de dispositifs lourds d'observation :

— dans ce cas, si la question posée est de nature descriptive, il se peut que ces contraintes imposent le recours à des études de cas, de préférence à des analyses basées sur des enquêtes statistiques qui nécessitent davantage de temps et de moyens ;

— si l'on s'interroge sur les liens de causalité entre la politique et les effets observés, de telles contraintes n'autorisent guère la mise en œuvre d'approches expérimentales ou quasi-expérimentales qui seraient par ailleurs les plus adaptées ; dans ce cas l'évaluation n'a pas lieu d'être entreprise ;

— le groupe a pu constater que les comparaisons de groupes témoins non équivalents issus de l'observation sont les plus fréquemment retenues en situation de pénurie relative de moyens. Ceci n'est pas le fruit du hasard : en dépit des critiques méthodologiques qui peuvent lui être adressées, ce plan de recherche réalise probablement le meilleur compromis entre le respect des contraintes et le souci de validité de l'évaluation. Si sa validité interne n'est en soi guère

satisfaisante, il permet toutefois de dégager des présomptions de causalité assez solides, surtout lorsqu'on peut multiplier les groupes témoins indépendants (ce fut notamment le cas pour la prophylaxie de la brucellose et à un moindre degré dans le cadre de l'évaluation du plan textile). A cet égard, pour l'évaluation de nombreuses politiques publiques, le recours à des comparaisons internationales ou régionales constitue sans doute un instrument précieux, mais à utiliser et surtout à interpréter avec prudence ;

— lorsque la comparaison entre groupes non équivalents n'est pas envisageable, l'analyse de séries temporelles (avant-après) peut constituer une solution de repli. Toutefois les corrélations qu'elle est susceptible de montrer entre des inflexions ou des ruptures de tendance d'une part et la mise en œuvre de la politique d'autre part ne prouvent pas vraiment l'existence de liens de causalité ;

— si l'analyse avant-après n'est pas non plus possible, il est nécessaire de se contenter d'analyses *ex post* reposant sur des enquêtes statistiques, sur des études de cas ou encore sur des avis d'experts. Cependant, ces techniques se limitent dans les cas les plus favorables à confirmer (ou plutôt à ne pas infirmer) l'existence d'un lien de causalité que l'on avait de fortes raisons de soupçonner au départ. Le recours à des avis d'experts peut également se révéler très utile à titre d'instrument de complément, de même que l'évaluation de synthèse consistant à tirer parti des études d'évaluation déjà réalisées.

• Les contraintes de délais risquent d'apparaître particulièrement **incompatibles avec le souci de généraliser les conclusions de l'évaluation**, à moins que des comparaisons de groupes non équivalents permettent de déterminer sous quelles conditions la politique évaluée peut contribuer efficacement aux objectifs poursuivis. Le recours simultané à plusieurs plans de recherche de mise en œuvre rapide, à l'image de ce qui a été réalisé dans l'évaluation du plan textile, peut également constituer un moyen de préserver la validité de l'évaluation en dépit de ces contraintes de délais.

III - 4 L'après-évaluation

Le rapport d'évaluation doit théoriquement fournir aux décideurs des éléments leur permettant de statuer en pleine connaissance de cause sur l'avenir de la politique étudiée : poursuivre cette politique, la renforcer, ou au contraire l'infléchir, voire y mettre fin. Ceci implique en premier lieu un effort de « lisibilité » de la part des auteurs du rapport. Ceci suppose en outre que l'évaluation possède une validité interne et externe suffisante et que les limites de cette validité soient précisées. Sinon, les résultats de l'évaluation risquent d'être utilisés de façon inappropriée, ou bien encore les divergences d'appréciation quant aux effets de la politique suivie ne seront pas vraiment atténués : ainsi, le bilan du

L'évaluation de la politique d'aide à l'élevage bovin (*)

L'élevage bovin se caractérise en France à la fois par son importance (près des deux tiers des exploitations agricoles sont concernées) et par les retards qui l'affectent (en particulier une productivité nettement inférieure à celle enregistrée en Allemagne et aux Pays-Bas), en dépit d'une longue politique de soutien des pouvoirs publics (de 1972 à 1983 l'effort financier de l'État est passé de 5 % à 14 % de la valeur de la production de la filière bovine).

La réalisation de l'étude prioritaire interministérielle d'évaluation confiée à Michel Colonna, ingénieur général du Génie rural et des Eaux et Forêts, s'appuie sur un groupe de travail composé pour l'essentiel d'ingénieurs généraux du Conseil général du Génie rural et des Eaux et Forêts (CGREF), de représentants du ministère de l'Agriculture et du ministère de l'Économie, des Finances et du Budget.

La démarche de l'étude d'évaluation :

1. Les niveaux d'objectifs considérés :

L'étude a distingué deux niveaux d'objectifs de la filière bovine : des **objectifs généraux** qui ne sont pas spécifiques à l'élevage bovin (contribution à la croissance économique et au commerce extérieur, maintien du cadre de vie rural, etc.), et des **objectifs opérationnels** (organisation de la filière bovine, amélioration de la productivité). A ces objectifs contribuent des actions regroupées en programmes (amélioration génétique, prophylaxie, interventions sur les marchés, etc. (cf. encadré n° 1, p. 30). Le travail du groupe a essentiellement porté sur les objectifs opérationnels.

Cette démarche n'a pas toujours pu être suivie, notamment lorsque coexistaient au sein d'un programme des actions d'orientation (associées à des objectifs explicites) et des actions d'accompagnement (destinées à répondre au coup par coup à certaines difficultés). C'est le cas notamment de la politique de modernisation des abattoirs publics.

2. Les niveaux d'efficacité distingués :

Trois niveaux d'évaluation de la politique d'élevage bovin ont été distingués :

- l'évaluation de la réalisation des actions ;
- l'évaluation du degré d'atteinte des objectifs ;
- l'évaluation de l'impact des mesures sur la satisfaction des bénéficiaires (à partir d'entretiens avec des acteurs locaux de la politique).

3. Les plans de recherche utilisés :

L'appréciation des effets de la politique a été réalisée essentiellement à partir de comparaisons de groupes témoins non équivalents :

(*) Voir la communication au groupe de Claude le Pen : présentation méthodologique de l'étude d'évaluation de la politique publique de l'élevage bovin.

Encadré n° 3 (suite)

- comparaisons internationales des performances de la filière bovine en France et chez nos principaux partenaires européens (Allemagne, Pays-Bas, Royaume-Uni) ;
- comparaisons régionales dans le cas de la prophylaxie de la brucellose (cf. encadré) ;
- comparaisons entre exploitants pour analyser les conséquences des plans de développement.

Lorsque ce plan de recherche n'a pu être mis en œuvre, le groupe a eu recours à des jugements d'experts (exemple : aide aux bâtiments d'élevage, monopole de zone dans le cas de l'insémination artificielle).

plan textile, mené par un groupe de travail interministériel a-t-il donné lieu à des interprétations différentes, en ce qui concerne l'efficacité du plan. La poursuite des travaux d'évaluation et le recours à des techniques et des plans de recherche plus performants peuvent, dans ce type de situation, être proposés comme un préalable aux décisions à prendre en ce qui concerne la politique analysée, surtout s'il s'agit de l'étendre à d'autres groupes. Toutefois, cette conclusion de l'évaluation en forme de poursuite du processus d'évaluation n'a pas lieu de servir de prétexte pour temporiser ou pour voiler le fait que la question posée n'appelle pas vraiment de réponse.

A l'image du contrôle administratif, **l'évaluation doit pouvoir donner lieu à un échange contradictoire** avec les services responsables de l'élaboration et de la mise en œuvre de la politique. Cela suppose que ces services disposent des éléments qui leur permettent de répondre point par point aux arguments avancés par le groupe d'évaluation (accès aux sources d'informations utilisées), et qu'ils soient à même de critiquer s'il y a lieu la méthodologie adoptée. Il se peut toutefois que cette dernière condition soit difficile à remplir lorsque l'évaluation repose sur des approches à caractère technique (la modélisation par exemple) à moins que ces services ne se dotent, avec l'appui de centres de recherches, d'instruments concurrents de même nature, mais reposant sur leur propre interprétation des mécanismes de la politique.

La **publicité** du rapport d'évaluation est un aspect sur lequel il n'est pas possible de fournir des recommandations systématiques, sauf à donner dans la naïveté. Ainsi aux États-Unis, le degré de diffusion des évaluations réalisées par le GAO est-il très varié. S'il est a priori souhaitable que les résultats soient, dans toute la mesure du possible, rendus publics, il apparaît légitime que la décision relève du **prescripteur** de l'évaluation. S'agissant de politiques très controversées, il se peut que l'éventualité d'une trop large diffusion incite les évaluateurs à une prudence exagérée. Enfin, certaines évaluations sont destinées par nature à une confidentialité plus marquée (celles notamment qui concernent des politiques touchant à la défense nationale). Une norme minimum peut consister à annoncer, avant l'évaluation, si celle-ci sera, ou non, publique.

Un schéma gradué peut également constituer une bonne solution dans certains cas. Il consisterait à séparer deux niveaux de publicité :

— un ensemble de données de bases déjà analysées, mises en perspective, mais sans qu'aucun jugement de valeur ne soit porté. Ce document serait partagé entre tous les partenaires (y compris le public) ;

— un rapport qui porte explicitement un jugement et qui peut dans certains cas rester confidentiel. Dans les deux cas le prescripteur dispose d'un droit d'antériorité qui lui permet de préparer sa « défense ».

III - 5 Pour une déontologie de l'évaluation des politiques publiques

A ce point de l'analyse il importe d'avancer les éléments d'un « code professionnel » propre à permettre l'exercice d'une évaluation conforme à ses finalités, en évitant les dangers précédemment mentionnés.

Ce code doit évidemment rester simple, eu égard aux variétés de situations concernées. Sa définition nous paraît pouvoir s'inspirer à la fois de certaines règles suivies en matière de contrôle administratif comme de production scientifique.

Deux grandes règles des contrôles administratifs semblent susceptibles d'avoir leur contrepartie dans le domaine de l'évaluation : le caractère contradictoire du contrôle, la séparation entre contrôleur et contrôlé. Plusieurs considérations justifient cette référence. Tout d'abord, mis à part les aspects plus spécifiquement techniques comme les finalités poursuivies, la conduite de l'évaluation soulève des questions dont certaines sont similaires à celles rencontrées par le contrôle administratif. On peut donc bénéficier de l'expérience acquise dans le cadre de cette activité. Les résultats en sont d'ailleurs probants. La mise en place des contrôles administratifs avait ainsi pour objet de garantir une administration « régulière »⁶. En dépit d'écarts ponctuels toujours fâcheux qui font les délices de la presse, on peut estimer, notamment par comparaison à la situation d'autres pays, que cet objectif a été globalement atteint en France.

L'absence de lien avec l'objet étudié, la confrontation avec des points de vue différents, sont des soucis qui se retrouvent également dans l'activité scientifique. Ils autorisent la production, le partage et l'accumulation des connaissances.

Nous proposons donc les deux règles fondamentales suivantes pour la conduite de l'évaluation :

Reproductibilité : l'évaluation doit être conduite suivant des protocoles qui la rendent reproductibles par d'autres évaluateurs.

6. Voir, sur ces points, l'introduction de l'ouvrage de Pierre Milloz, *Les inspections générales ministérielles dans l'administration française*, 1983, Éditions Economica.

Séparation : l'évaluateur, responsable des conclusions de l'évaluation, doit être une personne indépendante des agents (ou organisations) responsables de la formulation ou de l'exécution de la politique évaluée.

Chacune de ces règles, principalement la deuxième, moins généralement acceptée a priori que la première, appelle quelques commentaires :

1) La règle de **reproductibilité** correspond à l'impératif simple suivant : étant donné une évaluation réalisée par un certain évaluateur, tout autre évaluateur disposant des mêmes informations et suivant les mêmes procédures de recherche doit pouvoir aboutir au même résultat.

Cette règle passe en particulier par **l'affichage d'un protocole clair d'évaluation** précisant l'objet de l'évaluation, les sources d'information utilisées, les plans de recherche et les techniques employées. Cette exigence de transparence implique également le refus d'« arguments d'autorité » et constitue une garantie contre les risques de charlatanisme.

La fidélité à cette règle permet d'éclairer les utilisateurs de l'évaluation sur ses limites de validité. Elle autorise par ailleurs la comparaison d'évaluations différentes faites sur le même sujet.

2) La règle de **séparation** apparaît moins immédiate à certains. Dans sa discussion nous mettrons à part les pratiques consensuelles auxquelles le terme d'évaluation est parfois associé mais dans un sens différent de celui retenu ici et qui concernent plus des phases de formulation de politiques que de véritables appréciations *ex post*⁷.

Ce problème de séparation se pose évidemment dans le cas où une organisation a été impliquée de manière active dans la formulation ou l'exécution de la politique évaluée. Cette organisation peut alors avoir un avantage, ou un monopole, d'information. En outre, en particulier lorsque l'organisation concernée « produit » vraiment la politique évaluée (éducation, santé...), l'évaluation de cette politique peut être ressentie comme une évaluation de personnes. Dans ces conditions, convient-il, et si oui comment, d'impliquer l'organisation dans l'évaluation ?

Plusieurs justifications, générales ou particulières, d'une association entre évaluateurs et « évalués » ont été présentées. Ainsi D.T. Campbell (« Assesing... ») souligne le coût élevé qui peut être associé à une évaluation faite totalement à l'extérieur de l'organisation concernée. De même, P. Mondon souligne la nécessité d'une association étroite des enseignants au système d'observation mis en place par le ministère de l'Éducation nationale⁸. Mentionnons, dans un domaine voisin et au plan des pratiques, que dans certains ministères les fonctions d'inspection sont exercées par d'anciens « opérationnels »⁹.

7. Voir les exemples précédents des conférences-consensus proposées pour la formulation de recommandations de politique médicale ou des DRIR.

8. Ce système, à la fois transversal et longitudinal, est décrit dans l'encadré n° 3 de la partie II.

9. Il en est ainsi pour l'Agriculture et l'Équipement où les missions d'inspection sont traditionnellement confiées à des ingénieurs généraux (Génie rural et Eaux et Forêts, Ponts et Chaussées).

Il convient sans doute, en la matière, de ne pas avoir de position trop dogmatique. L'association des responsables de l'exécution de la politique de l'évaluation peut être nécessaire, à la fois pour préciser le passage des objectifs généraux aux objectifs opérationnels et pour éviter des réactions de rejet, soit du dispositif d'observation, soit des analyses de résultats. Au-delà de ces seuls objectifs, elle paraît présenter plus d'inconvénients que d'avantages, à la fois pour l'impartialité de l'évaluation et pour la diffusion de ses résultats. Reprenons ainsi le cas de l'éducation et supposons, scénario imaginé ici, qu'une étude fasse apparaître de meilleurs résultats pour les élèves du secondaire dont les professeurs proviennent de tel recrutement plutôt que de tel autre. Il est fort probable que, pour ne pas « désespérer » tel ou tel groupe d'enseignants, l'étude garde un caractère strictement confidentiel et ne reçoive pas certaines suites souhaitables (en matière de formation, de recrutement... ou de débat public !). L'auto-évaluation, si elle ne débouche pas nécessairement sur l'autosatisfaction, a par contre toute chance de conduire à l'autocensure, sauf à s'imposer des contraintes de publicité.

Au total, le principe de la séparation entre évaluation et exécution paraît devoir être affirmé. Sur ce point, nous rejoignons totalement la position prise dans le rapport sur l'évaluation des programmes de recherche¹⁰. Ces programmes doivent faire l'objet d'un suivi par des instances spécifiques, conseils de programme ou groupes d'évaluation et de prospective (GEP). Le rapport recommande que l'évaluation *ex post* soit confiée à des instances distinctes. Reprenons-en les considérations :

« Ce travail fait en continu (nota : de pilotage et de suivi par les GEP), outre son intérêt pour la conduite du programme, pour son bon déroulement, pour son adaptation continue, constituera la matière première indispensable de l'évaluation *ex post*. C'est pourquoi, la question a pu légitimement être posée de savoir s'il ne serait pas plus simple et plus efficace de déléguer totalement le processus d'évaluation à ces organes.

La réponse donnée est cependant négative : les GEP ou conseils de programme participent de façon trop directe à la conduite du programme pour pouvoir être investis d'une mission d'évaluation *ex post*.

Les hommes mêmes doivent être différents : il ne semblerait pas souhaitable de demander à un membre d'un GEP de faire partie, *intuitu personae*, d'un groupe d'évaluation. L'exigence d'un regard neuf, voire provocateur, se fait sentir, ce qui conduit à proposer de confier l'évaluation des programmes à des experts qui ne soient pas issus des instances du programme, mais qui devront travailler en étroite liaison avec elles... »

Mentionnons pour terminer que les justifications de l'évaluation en association, par des considération de « démocratie », ou l'appel à l'apprentissage qu'elle réalise, ne sont pas toujours dépourvues d'arrière-pensées. Elles peuvent, parfois, traduire le refus d'un regard extérieur, le souhait du maintien d'une structure close¹¹. Sur ce point, on peut citer le commentaire fait par R. Fraisse sur

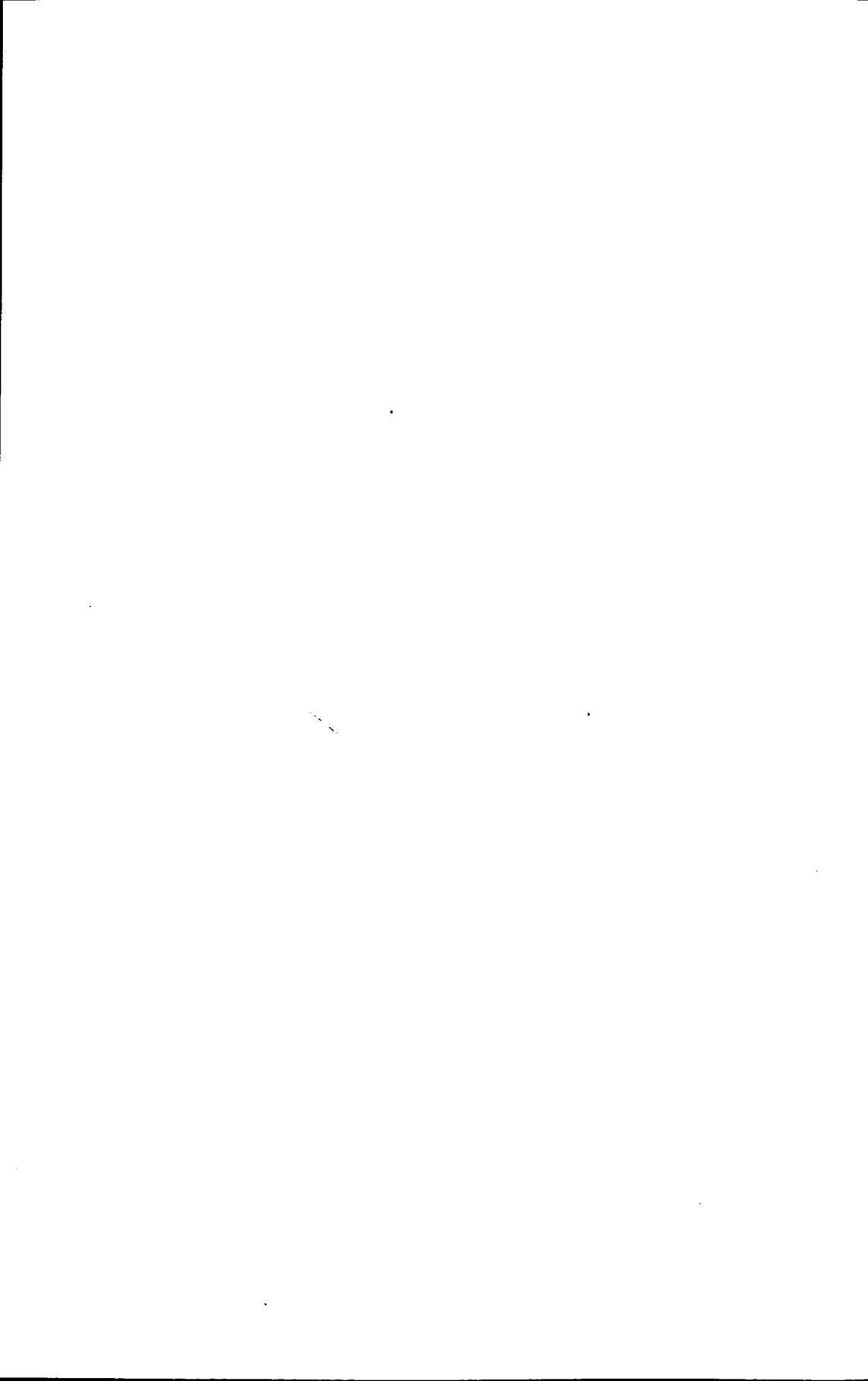
10. Voir en introduction p. 17.

11. Les partisans de l'auto-évaluation manifestent souvent une plus grande sollicitude pour l'organisation responsable de l'exécution de la politique publique que pour les citoyens qui en reçoivent les effets ou qui la financent.

l'évaluation-apprentissage¹², opposée à l'évaluation-sanction (*Compte-rendu du colloque : l'évaluation des politiques publiques*, Politique et Management Public, mars 1985).

12. Dans les termes du commentaire de R. Fraisse, l'évaluation-apprentissage concerne en fait l'auto-évaluation réalisée à des fins d'apprentissage.

« L'évaluation-apprentissage est trop facilement présentée comme la forme "progressive" de l'évaluation qui apprend à chacun à progresser en responsabilité et en autonomie. Cependant, il y a une utopie derrière la généralisation de l'évaluation-apprentissage : participant d'une conception naïve de l'autogestion, elle oublie volontiers l'existence et l'opinion des autres. Le dialogue avec des tiers est une nécessité : il doit pouvoir être un débat et le débat doit pouvoir accepter le conflit. Toute évaluation peut conduire à des ruptures dans l'action entreprise : la prise en compte de ces discontinuités appelle l'intervention des autres et la sanction de leurs jugements ainsi que la netteté de leurs avis, parfois brutaux parce qu'extérieurs. L'apprentissage, c'est bien aussi l'apprentissage du fait qu'on n'est pas seul... »



IV
Les domaines de l'évaluation



IV

Les domaines de l'évaluation

IV — 1 Un découpage des politiques publiques

Les développements précédents ont concerné la « boîte à outils » disponible pour réaliser une évaluation de politique publique (plans de recherche, techniques) et son « mode d'emploi » (mise en œuvre). Des exemples particuliers d'évaluation ont permis de les illustrer.

Au-delà de ces exemples, il importe de reconnaître si un découpage des politiques publiques par grands types ou catégories peut s'avérer utile vis-à-vis de la démarche évaluative. Plusieurs sont envisageables *a priori* : par secteur (agriculture, industrie, transports, etc.), par niveau (mesure élémentaire, programme...), en distinguant politiques ministérielles et interministérielles, etc.

Quelles qualités rechercher dans l'adoption d'un découpage particulier ? Tout d'abord, à des fins pédagogiques mais aussi opérationnelles, le découpage retenu doit pouvoir être mis en rapport avec les catégories conceptuelles, les nomenclatures, les sources existantes. Ces données peuvent présenter, pour tel cas particulier, des limites dont il convient de s'affranchir par des travaux spécifiques (enquête *ad hoc*, modélisation théorique, etc.). Il demeure que, pour obtenir un cadrage synthétique immédiatement utilisable, la référence à l'existant est nécessaire. Par ailleurs, compte tenu de l'objet du rapport, le découpage doit être discriminant vis-à-vis des aspects techniques et organisationnels de l'évaluation et permettre de dégager quelques conclusions globales simples, sur les plans de recherche, les techniques plus particulièrement adaptées (ou adaptables) aux diverses catégories de politiques définies.

Il serait vain de rechercher un découpage idéal excluant tout arbitraire ou redondance. Une assurance toutefois : un découpage sectoriel, de type ministériel, serait peu pertinent. En effet, la « spécificité » des politiques sectorielles ne doit pas être surestimée comme c'est fréquemment le cas¹. Ainsi, l'évaluation de l'efficacité (ou de l'inefficacité) d'une procédure d'aide aux entreprises ne soulève pas des

1. L'argument de spécificité est, presque systématiquement, opposé aux « outsiders » de tout poil, contrôleurs, analystes de politiques publiques, évaluateurs ou économètres, qui prétendent appliquer une méthodologie générale au domaine particulier considéré. S'il traduit, souvent, un simple réflexe de défense de pré carré, son usage peut également procéder d'une surestimation de l'importance des facteurs institutionnels... ou de la simple ignorance des autres domaines. L'examen, par exemple, de politiques sectorielles diverses (industrie, agriculture, transports...) tend plutôt à conforter la proposition suivant laquelle les mêmes causes produisent toujours les mêmes effets.

problèmes très différents, que ces entreprises appartiennent à l'agriculture ou à l'industrie manufacturière. Il en est de même pour les diverses réglementations conférant une rente à une profession particulière (taxis, pompes funèbres, transports aériens, etc.).

En définitive, il s'avère qu'un découpage traditionnel, inspiré de la théorie économique des finances publiques² constitue un choix adapté. On distinguera les domaines suivants : production de biens et services, interventions économiques financières (incitation, redistribution, taxation), réglementation, communication sociale, régulation macroéconomique.

Outre sa correspondance immédiate avec une conceptualisation existante de l'action publique, ce découpage présente plusieurs avantages.

Tout d'abord, pour les politiques publiques donnant lieu à des flux financiers enregistrés par des documents de comptabilité publique ou de comptabilité nationale, il permet un recouplement avec les nomenclatures comptables existantes³, soit en terme de catégories de dépenses, soit en termes de fonctions. Ainsi la ventilation du budget de l'État suivant la nomenclature fonctionnelle budgétaire fait immédiatement ressortir le pôle « production de biens et services », à fort contenu en main-d'œuvre (éducation, défense) et le pôle « incitation, redistribution », où, au contraire, les dépenses dites d'intervention publique sont dominantes (exemple : Sécurité sociale, santé, emploi). Le découpage retenu permet donc de délimiter quantitativement sans trop de difficultés les domaines d'action concernés. On reviendra sur ce point plus bas.

Un deuxième avantage du découpage retenu est d'être discriminant par rapport à l'acquis existant en matière d'évaluation ou d'approches connexes. Ainsi, le contrôle administratif a principalement concerné la production de biens et de services et la régularité de gestion des organisations publiques responsables. Il s'est beaucoup moins exercé sur les interventions économiques financières et, *a fortiori*, sur la réglementation⁴.

Enfin, cette approche satisfait à l'ambition initiale d'obtenir quelques conclusions globales sur les plans de recherche et les techniques utilisés ou utilisables pour chaque catégorie retenue. Le découpage proposé, inspiré *a priori* de la théorie des finances publiques, recoupe en effet des facteurs de différenciation décisifs pour le choix de plans de recherche : mise en œuvre d'une organisation publique complexe (production de biens et de services) ou pas (interventions économiques financières), niveau (régulation macroéconomique ou politiques micro), etc.

Au total, s'il demeure des difficultés liées au caractère non homogène des statistiques existantes⁵, à l'absence de comptes synthétiques (exemple : secteur hospitalier) ou, éventuellement, de données chiffrées (réglementation), le choix de catégories proposé permet de cerner, qualitativement et quantitativement, les

2. Voir par exemple les ouvrages de R. Musgrave.

3. Ce recouplement n'est pas fortuit : la nomenclature comptable s'inspire elle-même, en partie, de la théorie économique des finances publiques.

4. Ou, pour le moins, sa démarche interne laisse totalement dans l'ombre l'aspect essentiel dans ce cas, à savoir l'impact sur l'extérieur.

5. Ainsi, en matière de statistiques de comptabilité nationale, l'activité publique est retracée à la fois dans le compte des administrations, des grandes entreprises nationales, de l'industrie (pour le secteur public industriel), etc.

grands domaines d'intervention de l'État, et de manière pertinente pour l'évaluation.

En particulier, la référence au compte des administrations⁶ de la comptabilité nationale permet de préciser (ou de rappeler) plusieurs caractéristiques importantes sur le champ et l'organisation des interventions publiques que ces administrations (au sens de la comptabilité nationale) assument. On rappelle que ce compte ne retrace pas directement l'activité des entreprises publiques (y compris PTT) et du secteur hospitalier (sinon au travers d'opérations de répartition). Avec cette restriction en tête, les conclusions suivantes peuvent être obtenues (voir tableaux 1 et 2) :

— **l'essentiel des activités de production de biens et de services (non marchands) par les administrations est le fait des administrations centrales** (tableau 1). Appréciée au travers des salaires versés (ou éventuellement de la « consommation finale » au sens de la comptabilité nationale), la majeure partie correspond aux fonctions **d'éducation** (48 % des salaires versés par l'État) ainsi qu'aux **fonctions régaliennes** usuelles : défense (19,4 %), sécurité intérieure (11 %), administration financière (7 %).

Les autres fonctions de production de biens et de services sont assurées surtout à la **périphérie de l'État** (santé par exemple).

— pour chaque fonction, le **poids relatif des activités de production de biens et de services** et des activités **d'intervention économique** peut être apprécié en considérant le **rapport entre les dépenses** (au sens de la comptabilité nationale) et le **coût de production**. Un rapport élevé est en général caractéristique d'une importance marquée des « opérations de répartition ». Aux fonctions « de production » précédentes (enseignement, administration générale, justice et sécurité, défense nationale), pour lesquels le rapport dépense totale/coût de production est voisin de l'unité, s'opposent clairement les fonctions santé et interventions sociales (tableau 2).

— au plan global, la référence aux dépenses des administrations publiques au sens de la comptabilité nationale (tableau 2) montre que le **pôle « interventions économiques financières »** l'emporte largement sur le **pôle « production de biens et de services »**. Cette prédominance est essentiellement due au poids des administrations de **sécurité sociale** qui n'ont, en elles-mêmes, qu'un rôle de **tuyauterie redistributrice**⁷. La partie directement active (ou supposée telle) du pôle, correspondant par exemple aux diverses aides au secteur productif, est **plus modeste**. En l'identifiant aux « subventions d'exploitation versées » de la comptabilité nationale, elle représenterait environ 4,5 % des dépenses des administrations publiques. **L'essentiel de ces interventions « actives »** relève des **administrations publiques centrales**, les interventions redistributives étant surtout le fait, comme il a été dit, des administrations de sécurité sociale.

6. Plus précisément, à la présentation des dépenses et des coûts de production par fonction.

7. Avec, encore une fois, cette restriction que la partie production proprement dite (secteur hospitalier) est incluse, à ce niveau, dans une description en termes de transfert.

TABLEAU 1 : COÛTS DE PRODUCTION PAR FONCTION DES ADMINISTRATIONS PUBLIQUES
 (Année 1982 - Millions de francs courants - Comptes de la Nation)
 La répartition entre administrations est donnée entre parenthèses en pourcentage.
 Le pourcentage le plus important est en gras.

Administ. pub. centrales	Administ. pub. locales	Administ. pub. de Séc. soc.	Total	Fonctions
175 489 (87,4)	25 383 (12,6)		200 872	Enseignement, formation
5 119 (34,4)	9 781 (65,6)		14 900	Culture
2 241 (12,8)	2 883 (16,4)	12 436 70,8	17 560	Santé
4 450 (9,7)	14 742 (32,1)	26 684 58,2	45 876	Interventions sociales
1 013 (95,1)	52 (4,9)		1 065	Logement
2 158 (8,0)	24 660 (92,0)		26 818	Développement urbain
2 519 (29,9)	5 897 (70,1)		8 416	Développement rural
14 465 (68,3)	6 705 (31,7)		21 170	Transports, télécommunications
23 653 (71,5)	9 428 (28,5)		33 081	Action économique
39 587 (57,0)	29 840 (43,0)		69 427	Administration générale
9 086 (99,9)	7 (0,1)		9 093	Recherche
26 013 (83,2)	5 234 (16,8)		31 247	Justice, sécurité
4 718 (100,0)			4 718	Relations extérieures
404 (100,0)			404	DOM-TOM
148 275 (100,0)			148 275	Défense nationale
581 (60,1)	385 (39,9)		966	Non ventilé
459 771 (72,5)	134 997 (21,3)	39 120 (6,2)	633 888	TOTAL

Ce cadrage, pour incomplet qu'il soit, permet d'apprécier la surface des divers domaines de l'activité publique. Nous allons maintenant les passer en revue du point de vue de leur spécificité vis-à-vis des méthodes de l'évaluation.

**TABLEAU 2 — COÛTS DE PRODUCTION ET DÉPENSES TOTALES DES ADMINISTRATIONS
PUBLIQUES PAR FONCTION**

(source : rapport sur les Comptes de la Nation de l'année 1984 — INSEE Série C 124-125)
(Année 1982, Milliards de francs constants)

Opérations Administrations	Consommations inter- médiaires (1)	Rémunération salariés (2)	Impôts nets (3)	Coût de production (4) = (1)+(2)+(3)	Consommation finale (5)	Opérations courantes de répartition (6)	Opérations en capital (7)	Dépenses totales (8) = (5)+(6)+(7)
Enseignement, formation	34 437	163 583	2 852	200 872	189 388	9 994	13 567	212 949
Culture	5 972	8 805	123	14 900	21 366	6 988	11 663	40 017
Santé	3 639	13 155	766	17 560	17 532	215 413	2 751	235 696
Interventions sociales	13 552	30 858	1 204	45 614	42 697	676 010	6 303	725 010
Logement	720	339	6	1 065	1 247	37 533	8 236	47 016
Développement urbain	11 438	16 080	- 700	26 818	23 134	14 876	28 105	66 115
Développement rural	3 870	4 625	- 79	8 416	7 970	3 420	9 383	20 773
Transports, télécommunications	10 584	10 892	- 306	21 170	14 868	15 626	18 939	49 433
Action économique	15 603	16 481	997	33 081	26 461	40 784	15 172	82 417
Administration générale	19 490	49 509	690	69 689	60 746	1 724	11 323	73 793
Recherche	2 529	5 320	1 244	9 093	9 081	2 869	1 244	13 194
Justice, sécurité	3 934	26 790	523	31 247	32 470	621	2 197	35 288
Relations extérieures	816	3 833	69	4 718	4 764	7 375	430	12 569
DOM-TOM	154	245	5	404	419	9 493	1 134	11 046
Défense nationale	73 760	72 862	1 653	148 275	124 552	5 283	6 943	136 778
Non ventilé	563	400	3	966	1 794	68 340	2 550	72 684
TOTAL	201 061	423 777	9 050	633 888	478 489	1 116 349	139 940	1 834 778

IV - 2 La production de biens et de services

Les activités publiques correspondantes reposent sur la mise en œuvre de facteurs (travail, consommations intermédiaires, capital) au travers d'une organisation spécifique (entité administrative, établissement public, entreprise). Elles conduisent à fournir certains biens ou services à des « consommateurs » externes, publics ou privés.

Les activités couvertes par cette description présentent à la fois des caractéristiques communes fondamentales et des caractéristiques de différenciation.

Première caractéristique commune : le fait que les politiques publiques associées soient le fait d'organisations complexes faisant appel à une « main d'œuvre » abondante. Il en résulte que l'analyse « interne » de l'organisation peut être décisive pour l'évaluation. Le problème sera généralement différent pour les politiques incitatives ou réglementaires. Ce poids des aspects « organisation » est fortement discriminant pour les plans de recherche envisageables et, en particulier, dans bon nombre de cas, limite le recours à des expérimentations « vraies ».

Deuxième caractéristique commune, corollaire de la première : l'existence d'une forte tradition de contrôle administratif, essentiellement de régularité, fondée sur l'exploitation de documents comptables (voir I.3.).

On notera que cette tradition concerne surtout l'organisation de la production elle-même et que l'analyse de l'impact externe, qui se développe actuellement dans les contrôles administratifs, demeure encore limitée.

La troisième caractéristique générale tient également à l'existence d'une organisation lourde qui donne un poids particulier à la dimension temporelle. Il en résulte une continuité entre les opérations de suivi de résultats, stricto sensu, et des opérations d'évaluation de politiques publiques consécutives à l'introduction d'un changement finalisé (réforme de l'enseignement, nouvelle tarification, extension du secteur public, etc.). Cette continuité est très sensible au plan des méthodes. On notera en particulier l'importance générale d'un recours à des dispositifs de suivi statistique lourds, de champ cohérent entre périodes.

Mentionnons enfin deux autres caractéristiques, de différenciation.

L'une concerne la distinction marchand/non marchand. Elle renvoie à une difficulté spécifique aux activités non marchandes, à l'amont de l'évaluation proprement dite : celle de l'identification même du bien ou du service produit et de sa valorisation éventuelle.

L'autre caractéristique de différenciation concerne l'importance de l'interaction entre l'interne (organisation) et l'externe (milieu récipiendaire). Cette interaction peut jouer un rôle fondamental et complexe dans le cas de certains services non marchands, tel que l'éducation : les résultats obtenus dépendent à la fois des performances de l'organisation mais aussi des caractéristiques des bénéficiaires (exemple : paramètres socio-culturels). On conçoit que dans ce cas, le problème d'évaluation puisse être fortement multidimensionnel, d'où des difficultés pour réaliser de « vraies » expériences contrôlées.

1) La production « publique » de biens et services marchands couvre, en particulier, l'ensemble du secteur public productif (grandes entreprises nationales, secteur public industriel), les établissements publics à caractère industriel et commercial, certaines activités tarifées exercées par des administrations produisant essentiellement dans le non marchand.

Les questions d'évaluation (de résultats comme de politiques) les plus fréquentes concernent :

- l'évolution des résultats comptables ;
- les décisions sur les facteurs de production (emploi, capital) ou leur gestion ;
- les politiques de prix (tarifs en particulier).

Ces questions peuvent être étroites (évaluation de l'introduction d'une nouvelle forme de tarification) ou larges (évaluation du programme électronucléaire) ou même renvoyer aux statuts des organisations considérées (avantages liés à l'existence d'un monopole public).

On rappelle qu'il existe en la matière une tradition importante de contrôle administratif et de suivi. Les travaux correspondants sont parfois peu centrés sur l'efficacité proprement dite et comportent souvent des lacunes. Par exemple, il est rare que, dans l'évaluation des résultats de telle entreprise publique, le partage entre impact de la politique proprement dite de l'entreprise (efforts de productivité, choix d'investissement, choix de produits) et évolution du marché soit très clairement opéré.

L'acquis apparaît toutefois solide. Des directions de progrès prioritaires concernent l'amélioration des éléments de comparaison intertemporelle et externe et une meilleure séparation de l'effet des divers facteurs.

La première direction concerne à la fois les entreprises publiques « concurrentielles » et « non concurrentielles ».

Il conviendrait, tout d'abord, de disposer d'informations statistiques homogènes dans le temps, permettant de pratiquer de manière fiable un suivi ou des approches de type avant-après. L'expérience prouve aussi l'intérêt de comparaisons systématiques avec divers « groupes témoins », constitués par exemple d'entreprises intervenant dans des secteurs similaires en France et à l'étranger. Si ces groupes témoins ne peuvent pas être considérés comme totalement équivalents, de telles comparaisons sont très utiles pour répondre à des questions du type suivant :

— on observe une amélioration sensible des résultats de certains groupes chimiques publics : cette amélioration est-elle supérieure ou inférieure à celle des groupes chimiques privés ou étrangers ? Quelle est la signification des différences observées ?

— les coûts de production d'Air France sont-ils, ou non, supérieurs à ceux de ses concurrents étrangers ? si oui, pourquoi (productivité, rémunérations) ?

— dans quelle mesure l'effort électronucléaire de la France a-t-il bénéficié aux divers agents économiques nationaux (comparaison des tarifs industriels ou domestiques avec ceux pratiqués à l'étranger) ?

— etc.

L'existence de « groupes témoins » externes et le recours à des comparaisons intertemporelles fiables sont particulièrement intéressants pour les questions d'évaluation concernant la politique suivie en matière de facteurs de production (emploi, capital) ou de tarifs par les groupes publics en situation de monopole. En l'absence de concurrence par le marché, les risques de dérive sont notables.

La seule manière d'en apprécier la réalité est souvent de recourir à de telles comparaisons externes ou intertemporelles. Une problématique similaire s'applique à l'évaluation *ex post* des grands projets, à l'instar de l'analyse du TGV Sud-Est faite par la Cour des comptes.

Notre deuxième remarque concerne la prise en compte de l'environnement de l'entreprise. Elle est commune aux entreprises publiques « concurrentielles » et « non concurrentielles » pour ce qui est de l'évolution du marché : l'obtention par une entreprise publique de résultats corrects sur un marché conjoncturellement porteur n'est pas en soit un signe de réussite exceptionnelle. Elle se renforce dans le cas d'entreprises publiques fortement « régulées » par l'État, c'est-à-dire essentiellement les grandes entreprises nationales. L'intervention de l'État, soit au travers de contraintes de service public, soit par le biais de mesures de régulation macroéconomique ou industrielle, est un facteur à prendre en compte dans l'évaluation. Ainsi une évaluation de l'évolution des tarifs énergétiques ou téléphoniques ne saurait ignorer l'impact de la politique des prix ou les choix de financement de la filière électronique, le programme d'équipement nucléaire d'EDF répond à certains objectifs de lissage industriel, etc. Dans bien des cas, l'évaluation de l'entreprise publique se mue en évaluation du couple État tutelle-entreprise.

Au total, l'analyse rapide précédente suggère que pour le secteur public marchand, les progrès de l'évaluation des résultats et des politiques publiques concernent, de manière somme toute classique :

- la mise en place de systèmes d'information permanents permettant d'établir des comparaisons externes fiables, dans le temps ou dans l'espace ;
- une meilleure prise en compte des divers facteurs d'environnement (évolution des marchés, rôle de l'État tutelle) pesant sur les résultats.

Ces préalables satisfaits, les plans de recherche correspondants sont évidemment de type « non expérimental » (au sens de l'expérimentation vraie de la partie II).

A cet égard, les possibilités offertes par l'expérimentation vraie paraissent limitées du fait du poids des aspects organisationnels. Cette position a été très clairement exprimée par A. Rivlin⁸, qui souligne le caractère coûteux, ou impraticable, d'un recours à l'expérimentation, pour identifier des « fonctions de production » d'organisations. Lorsque trop de paramètres indépendants sont en cause, sans possibilité de moyenne statistique, la rigueur de l'expérience de laboratoire contrôlée, paraît difficile à atteindre. Il y a même un danger à s'appuyer sur des pseudo-expériences » dont les conclusions sont appliquées sans discernement sous couvert de vertus apparentes de rigueur. Cette objection à une pratique abusive de l'expérimentation se retrouvera avec une acuité accrue dans le cas d'organisations

8. Voir A. Rivlin (1974), *How can experiments be more useful ?* (op. cité)

« non marchandes » en forte interaction, pour leurs résultats, avec le milieu social, telle l'éducation.

Il demeure que pour des cas simples de politiques très élémentaires, modifiant un signal de l'entreprise à destination de l'extérieur (tarifs, qualité, nouveaux services...), l'expérimentation retrouve ses droits. Ainsi la mise en place sur l'ensemble du territoire de la tarification EJP (effacement jour de pointe) par EDF a-t-elle été précédée de quasi-expérimentations locales. De véritables expériences ont été pratiquées dans ce domaine à plusieurs reprises aux États-Unis⁹.

Rappelons enfin l'importance fréquente pour une évaluation portant sur des caractéristiques relativement globales (exemple : productivité du travail) d'avoir auparavant écrit une « histoire » suffisamment solide de l'entreprise concernée, les choix passés ayant souvent un poids déterminant.

2) Nous en venons maintenant aux politiques publiques de **production de biens et services non marchands**.

Leur évaluation soulève, pour partie, des problèmes identiques à ceux rencontrés dans le cas de la production publique marchande. L'existence d'une organisation complexe, support de la production et, par là, des politiques, implique que l'évaluation requiert des appareils de suivi statistique permettant des comparaisons intertemporelles ou entre institutions similaires.

Des difficultés supplémentaires et conjuguées affectent l'élaboration même de la procédure d'évaluation pour :

— **identifier le bien ou service produit ;**

— **identifier le « résultat » qui peut impliquer une interaction forte avec le milieu extérieur « consommateur » du bien ou du service.**

Nous ferons quelques observations succinctes sur chacun de ces points.

La première difficulté est bien connue. Un moindre mal consiste à définir des **indicateurs**, éventuellement multiples et apportant des éclairages complémentaires : soins délivrés par catégorie, pourcentage de diplômés, heures d'enseignement, publications, brevets, pourcentage d'affaires criminelles résolues, etc. On n'examinera pas davantage cette question du choix des indicateurs. Elle a fait l'objet d'une littérature abondante dans le passé. Notons que la seule utilisation d'indicateurs d'activité peut éclairer un diagnostic sur l'efficacité des organisations concernées mais non sur leur efficacité sociale. Rappelons aussi que la sélection *a priori* d'indicateurs peut, dans certains cas, induire des comportements qui nécessitent un redressement lors de l'évaluation (cf. politique anti-crime du gouvernement Nixon).

Le deuxième point important concerne la **dépendance des résultats par rapport à l'environnement extérieur**, et notamment aux « récepteurs ». Cette situation est caractéristique d'un grand nombre d'activités publiques non marchandes, par

9. On trouvera dans l'article de D.J. Aigner et E.E. Leamer, *Estimation of time of-use pricing response in the absence of experimental data : an application of the methodology of data transferability* (a. paraître dans *The Energy Journal*), de nombreuses références sur de telles expérimentations tarifaires et une description de certaines d'entre elles (Southern California Edison Company, Los Angeles Department of Water and Power, Kansas City Power and Light).

exemple, l'éducation, la santé ou, à un degré moindre, la sécurité intérieure. On connaît la masse de travaux sur les liens entre échec scolaire et facteurs socio-culturels. De même, l'évolution, à un plan global, des taux de morbidité dépend de facteurs non directement sanitaires, etc. Ces diverses difficultés ont été illustrées dans les parties II et III.

Ces deux remarques ont une implication directe pour le choix des procédures. Le diagnostic renforce en un sens celui opéré dans le cas marchand.

Une première conclusion est négative : dans le cas où les politiques publiques associées mettent en jeu de manière complexe l'organisation « productrice » et le milieu récepteur, le recours à l'expérimentation vraie doit être envisagé avec prudence. L'éducation est un exemple particulièrement significatif : le nombre de « paramètres » concernés, la longueur du processus éducatif, son caractère cumulatif rendent illusoire une reproduction des conditions d'une expérience contrôlée en laboratoire. L'échec du programme américain d'évaluation mené dans les années 1970 fondé sur une approche systématique en terme de « planned variations » (on fait varier un à un les paramètres supposés importants : taille de la classe, etc.) est significatif¹⁰. Exemple actuel, la déconsidération dont souffre la recherche pédagogique tient sans doute aux déboires rencontrés dans l'introduction, parfois tonitruante, et chaque fois, bien sûr, assise sur le résultat d'expériences indubitables, de réformes pédagogiques successives et paraissant contradictoires : méthodes d'enseignement des langues, des mathématiques, de la lecture ; classes de transition puis soutien ; disciplines d'éveil, etc. En l'occurrence, la conjonction de références abusives à des expériences ponctuelles de portée limitée et d'applications massives n'a pas été sans coût. **La prudence s'impose donc : rien n'est pire en effet, dans des domaines fondamentaux tels que l'éducation ou la santé, que des pseudo-certitudes assises sur des pseudo-expériences non généralisables.** Par exemple, les résultats observés dans des lycées-pilotes où la population des enseignants et des élèves, les moyens disponibles, sont très spécifiques, ne peuvent être extrapolés sans danger à l'ensemble de la population scolaire. De même, l'introduction d'une modification globale du système de gestion des hôpitaux, sans identification préalable de leurs caractéristiques spécifiques, ne peut être que faiblement conclusive. Dans ces domaines, plutôt que de viser l'expérience contrôlée décisive (et, par là, la politique miracle), il semble préférable d'avoir une approche d'accumulation progressive de connaissances de type « muddling through ». **D'où l'intérêt de systèmes de suivi long du type de celui géré par le Service de la prévision, des statistiques et de l'évaluation du ministère de l'Éducation Nationale. D'où l'intérêt également d'expérimentations locales, ne prétendant pas au rang d'expériences décisives, mais dont les résultats puissent être analysés et confrontés de manière pragmatique et conduisent à cette accumulation de variance chère aux statisticiens.**

Dans le cas d'activités non marchandes, la démarche expérimentale retrouve ses droits dès lors que s'efface la double complexité d'un processus de production multidimensionnel et de fortes interactions avec l'extérieur. Il en est ainsi pour l'évaluation de politiques sanitaires relativement simples où l'extérieur constitue

10. Voir A. Rivlin, *How can experiments be more useful ?* (op. cité).

un milieu largement « passif » : fluorisation des eaux, campagne de vaccination, etc. Dès lors que ces politiques cessent d'être strictement « biochimiques », l'expérimentation demeure valide mais les difficultés de conception et d'interprétation sont accrues¹¹.

Une dernière remarque. On notera que les fonctions régaliennes les plus traditionnelles n'ont fait l'objet, en France, que d'un nombre très restreint d'évaluations. Il est ainsi pour l'administration générale¹², la justice et, bien évidemment, la défense. Comparée à d'autres pays, cette préservation d'un sanctuaire peut être, éventuellement, considéré comme archaïque. Aux États-Unis par exemple, les études d'évaluation sur la police ou la justice sont légion. Le GAO n'est restreint dans ses pouvoirs d'investigation que vis-à-vis de deux organismes : la Central Intelligence Agency et le Federal Reserve Board. On peut ainsi mentionner qu'une évaluation (de formulation) de politique de défense, portant sur le programme d'armes chimiques envisagé par le Pentagone et réalisée par le GAO à la demande du Congrès, a conduit au rejet de ce programme¹³.

IV - 3 Les interventions économiques financières : incitations, redistribution, taxation

Nous en venons maintenant au domaine principal de l'action publique en termes de flux monétaires. Le rapide cadrage précédent, établi à partir des données de comptabilité nationale, montre au sein de cette catégorie l'importance des administrations de sécurité sociale.

Par opposition avec la production de biens et de services et dans une perspective d'évaluation, une caractéristique notable des activités publiques afférentes est, en général, la (relative) simplicité de « l'interne »¹⁴ : les organisations concernées ont essentiellement un rôle d'application de textes ou de gestion administrative et comptable. Il n'en résulte pas qu'elles n'appellent ni audits ni évaluations (tout au contraire !). Les questions soulevées, de régularité (application des textes...) ou d'efficacité (délais de traitement des dossiers...), sont, toutefois, en général plus simples que pour des administrations d'éducation ou de santé.

11. Voir par exemple les difficultés mentionnées dans l'article de J.E. Harris, *Prenatal medical care and infant mortality* (1982, in V.R. Fuchs ed, *Economic Aspects of Health*, NBER, The University of Chicago Press, Chicago, Il).

12. On rappellera que les administrations centrales sont largement préservées des désagréments des contrôles.

13. Voir la communication d'E. Chelmsky au colloque sur le management public. (26-28 septembre 1984) sur le thème « Program evaluation in the United States : what influence does the congressional use of evaluation findings have on public management at the federal level ? »

14. Nous insistons sur les termes « en général » et « relatif ». Il va de soi que l'application de procédure d'aides soulevant des problèmes de reconnaissance d'éligibilité ou d'appréciation économique peut être délicate. Il en est ainsi notamment des politiques d'incitation à l'innovation. Dans ce cas, une étude du fonctionnement interne retrouve évidemment ses droits.

Les coûts de production (salaires, frais de fonctionnement etc.) engagés directement pour les interventions économiques apparaissent donc comme des coûts de mise en œuvre. Les politiques publiques affectent, soit le revenu des agents concernés, soit les prix (ou des variables équivalentes) auxquels ils sont confrontés : les « résultats » dépendront donc, au premier chef, des réactions des agents à ces signaux financiers. Cette description, valide en première approximation, explique le recours, largement développé dans certains pays mais non en France, à des modélisations microéconomiques des comportements pour l'évaluation. On reviendra sur ce point plus bas.

Rappelons que les politiques publiques concernées peuvent affecter les dépenses de l'État (il en est ainsi pour les diverses prestations sociales ou les aides versées aux entreprises¹⁵), ou ses recettes (la variation des taux d'imposition ou des assiettes est un instrument d'orientation éventuelle des décisions des agents). De manière générale, une politique de soutien à un secteur ou à un groupe d'agents particulier pourra comporter des actions pesant à la fois sur les dépenses (aides) et sur les recettes (fiscalités dérogatoires, parfois baptisées « dépenses fiscales »).

Enfin, nous avons retenu la distinction traditionnelle entre interventions incitatives et redistributives. Cette distinction permet dans certains cas de séparer des configurations extrêmes. Elle recouvre le plus souvent une distinction de finalités ou de résultats : la plupart des interventions économiques auront à la fois des impacts incitatifs et redistributifs (avec des chassés-croisés involontaires).

Des exemples d'évaluation de ces politiques ont été discutés précédemment (aides à l'emploi, aides à l'élevage bovin etc.). La tendance récente est à leur accroissement. Malgré ce progrès, l'acquis demeure limité. Certes des cadrages quantitatifs existent. On peut citer les travaux effectués au début des années soixante-dix sur les transferts État-industrie¹⁶ ou les rapports sur les aides annexés à la loi de finances. Pour nécessaires qu'ils soient, ces constats demeurent surtout comptables et n'autorisent aucune conclusion sur l'efficacité des dispositifs. De même les contrôles administratifs peuvent mettre en évidence des irrégularités graves d'application et permettent alors, de conduire à l'inutilité des interventions. En l'absence de telles irrégularités, quand « tout marche bien », ils n'autorisent pas en général un jugement définitif d'efficacité économique.

On peut donc estimer qu'il y a, dans ce domaine des interventions économiques financières, une carence importante en matière d'évaluation. Il existe pourtant de solides présomptions de l'inefficacité incitative de nombreuses interventions économiques présentées comme des leviers d'action importants pour l'État¹⁷. Il importe de rendre plus décisives ces présomptions, ou de les invalider. Dans le système actuel, on peut s'interroger sur l'efficacité incitative de nombreuses aides aux entreprises, compte tenu des « degrés de liberté » des systèmes auxquels elles s'appliquent (une aide à l'investissement peut fort bien se muer en distribution de dividendes). Symétriquement, de nombreuses aides sociales, ou la fiscalité, souvent appréciées dans une perspective redistributive, génèrent sans aucun doute

15. Sous de multiples formes on le sait : aides sectorielles, bonifications d'intérêt etc.

16. A. le Pors Les transferts État-Industrie Statistiques et Etudes financières, série orange, n° 2, 1971.

17. Divers rapports administratifs ont été ainsi consacrés à certaines catégories d'aides publiques.

des effets incitatifs non voulus¹⁸. Telles sont des questions auxquelles une évaluation plus méthodique permettrait d'apporter une réponse.

Ce constat suggère trois voies de progrès.

La première consiste en un développement du **recours aux méthodes d'expérimentation**. Ces méthodes paraissent convenir aux cas d'interventions économiques financières s'adressant à des groupes nombreux (aides aux ménages en particulier, aides générales aux entreprises) et où les phénomènes de « cannibalisme »¹⁹ sont limités. Le développement de telles pratiques suppose toutefois une remise en cause de certaines traditions d'égalitarisme formel.

On évoquera ensuite **l'association, à tout nouveau système d'intervention économique** (aides, plan sectoriel, etc.), **d'un dispositif d'évaluation prévoyant, en particulier, un bilan d'efficacité à horizon donné et une remise en cause éventuelle de la politique**. Ce mécanisme²⁰, permettrait d'éviter l'empilement de systèmes d'intervention reconduits sous la double pression des groupes bénéficiaires et des administrations gestionnaires.

Enfin, l'évaluation des interventions économiques gagnerait souvent à s'appuyer sur une **modélisation microéconomique des comportements** avec estimation sur **données individuelles**. On a mentionné la maigreur des travaux réalisés en France dans ce domaine²¹. Le manque de jugements relativement assurés soit sur l'efficacité incitative de telle aide aux entreprises, soit sur l'absence d'effets pervers liés à un système particulier de prestations sociales ou de taxes découle, notamment pour les politiques existantes où l'expérimentation n'est plus possible, **du caractère le plus souvent comptable et non explicatif des constats**.

IV - 4 La réglementation

Du point de vue de l'évaluation, les **politiques de nature réglementaires ont des caractéristiques proches de celles reposant sur des interventions économiques financières** : relative simplicité des organisations administratives responsables²², dépendance quasi exclusive des résultats par rapport à la réponse, en termes de

18. De nombreux travaux anglo-saxons, procédant par expérimentation ou modélisation, ont été consacrés au thème de l'impact sur l'offre du travail d'aides sociales ou de taxes. Voir les références des actions publiques en matière d'emploi (1975-1985) », (communication aux Journées d'études des 3 et 4 octobre 1985, organisées par l'ANPE, le CGP et la délégation à l'emploi, ronéo).

Un exemple pour le cas de la France (avec une présentation générale et de nombreuses références) est fourni par l'étude de F. Bourguignon et B. Lemmenier *Fiscalité, para-fiscalité et offre de travail féminin* (1984, CREDOC, ronéo).

19. Dans le cas d'une procédure d'aide publique aux entreprises, ce terme désigne l'élimination d'entreprise non aidées par les entreprises aidées. A marché donné, de tels phénomènes peuvent rendre délicat l'extension à un plan global de conclusions éventuellement positives fondées sur la comparaison entreprises aidées — entreprises non aidées.

20. La « sunset legislation » aux États-Unis en fournit un exemple.

21. Voir les quelques études citées précédemment.

22. On fera les mêmes remarques que précédemment. Ce jugement sur la simplicité de l'interne est vrai à condition que la réglementation soit elle-même « relativement simple » à appliquer, ce qui n'est pas toujours le cas.

comportement, des agents concernés (et donc possibilité de détour et de neutralisation pour les systèmes récepteurs complexes). Les mêmes recommandations peuvent donc être formulées : développement de l'expérimentation, inclusion de procédures d'évaluation (avec possibilité d'extinction de la réglementation) dès la mise en place de nouveaux dispositifs réglementaires, recours à la modélisation microéconomique et à l'estimation sur données individuelles. Nous ne reviendrons pas sur ces divers thèmes.

Par rapport aux politiques précédentes, la réglementation se distingue toutefois par une caractéristique importante : l'absence fréquente de « trace » chiffrée générée directement par la mise en œuvre. Ainsi, le volume des aides à l'exportation est enregistré en comptabilité publique. Compte tenu de l'existence de dossiers d'aide, il peut être rapporté aux flux d'exportations concernés, mis en parallèle avec la situation des entreprises bénéficiaires, etc. Un premier cadrage quantitatif est donc réalisable à partir des données existantes. A l'opposé, l'évaluation des effets de la loi sur le prix unique du livre nécessitera des données spécifiques non générées automatiquement par l'application du dispositif réglementaire lui-même : évolution des chiffres d'affaires des libraires, des éditeurs, des ventes (en volume), du prix des livres, etc. Sauf mise en place d'un dispositif *ad hoc* de collecte, l'évaluation est impossible. Une telle situation est fréquente et l'acquis d'information en matière de politiques réglementaires est très en deçà de celui existant pour les interventions économiques.

Cette absence d'acquis concerne aussi les outils d'analyse conceptuelle comme les enseignements à tirer des contrôles administratifs. S'il existe des théories économiques ou politiques de la réglementation, les travaux réalisés dans ce domaine, notamment de nature appliquée, sont peu nombreux en France. Les enseignements de contrôles administratifs de régularité ne peuvent être entièrement conclusifs. Vérifier qu'une réglementation est appliquée est une chose. En apprécier l'impact en est une autre de nature très différente.

Au total, si l'analyse s'est développée dans ce domaine au cours des années récentes, la réglementation demeure un grand continent vierge de l'évaluation des politiques publiques.

L'accent devrait donc être fortement mis au cours des années à venir sur l'évaluation des politiques réglementaires, de manière d'autant plus volontariste qu'elles jouissent d'une immunité certaine. Ressenties comme moins directement coûteuses en termes budgétaires, elles ne suscitent pas les mêmes réserves *a priori* que des interventions économiques financières. Par ailleurs, les réglementations conduisant à des transferts entre agents n'ont de chance d'être remises en cause que si des groupes de pression suffisamment forts se mobilisent contre le groupe de pression bénéficiaire de la réglementation²³. Une telle situation s'observe parfois dans le cas de face à face entre producteurs également organisés. Les chances de remise en cause sont beaucoup plus faibles lorsque la réglementation transfère un surplus perçu sur des usagers dispersés et nombreux vers un secteur bien structuré. Il en est ainsi dans le cas des monopoles publics ou de certaines professions

23. Il y a là une des difficultés majeures en matière de « déréglementation ». Voir par exemple E.L. Thurow *The zero sum society*.

fermées²⁴. L'évaluation des politiques publiques afférentes est particulièrement justifiée pour de telles configurations. Une déréglementation partielle peut d'ailleurs constituer une approche expérimentale intéressante, dans certains cas (exemple : transport aériens).

Mentionnons enfin un dernier domaine où le développement d'une évaluation par expérimentations locales, pourrait présenter des avantages : celui de la **réglementation du travail**²⁵.

IV - 5 La communication sociale

Nous en venons maintenant à un domaine particulièrement immatériel des politiques publiques : celui de **l'action sur l'opinion**. Ce domaine concerne, bien sûr, l'ensemble des politiques publiques précédemment discutées. Il comprend aussi des actions spécifiques, les campagnes d'information de la puissance publique, pour lesquelles une évaluation systématique découle d'obligations réglementaires. Une double raison pour consacrer un court exposé à ce thème.

Sur le premier point, il va de soi que le contenu des politiques publiques en « communication sociale » peut être important. A un extrême, on trouve, dans les catégories précédentes, de multiples exemples de politiques symboles, coups médiatiques ou réponses à une émotion de l'opinion publique, dont les effets autres que de communication sont restreints. Une vision extrême est de considérer que toute politique publique **n'est que communication**. Nous n'adhérerons pas, dans ce rapport, à une telle vision, inexacte, sans nier que la référence à des objectifs de communication sociale puisse être fondamentale pour comprendre la **genèse** d'une telle politique particulière ou son **exécution**, et par là, constituer un aspect nécessaire de son évaluation.

Les campagnes publiques d'information proprement dites se sont développées de manière importante au cours des dix dernières années. On sait qu'une appréciation de leur impact est obligatoire. Il semblerait donc que la question de leur évaluation soit largement résolue. En fait, bien des difficultés demeurent²⁶.

Les pratiques actuelles restent en effet assez largement centrées sur la communication proprement dite. C'est **l'impact d'information** de la campagne qui est mesuré au travers d'indicateurs caractéristiques de la réception, de la

24. Des réglementations *a priori* moins conflictuelles, telles que les réglementations techniques, peuvent être également susceptibles de cette démarche d'évaluation. Notons qu'il peut également y avoir dans ce cas des transferts entre agents non nécessairement justifiés. Ainsi le choix des normes d'isolation thermique peut constituer un enjeu entre producteurs d'énergie, constructeurs, experts en économies d'énergie, etc.

25. Certaines difficultés rencontrées dans l'application des dispositions sur la réduction de la durée hebdomadaire de travail tiennent ainsi à une approche excessivement globale et ignorant le comportement des micro-agents.

26. Dans ce qui suit, on s'inspire largement de la communication faite par B. Dubois, Y. Evrard, D. Lindon, « L'évaluation des campagnes d'intérêt public » au séminaire organisé par l'ESSEC en 1980 sur le thème « L'État et la publicité » (note CESA, ronéo).

mémorisation ou de la compréhension. Par contre les effets sur les attitudes ou les comportements sont peu explorés. Les difficultés sont réelles. S'il est possible d'apprécier, par exemple, les modifications de comportement induites par des campagnes d'information routière (comptages statistiques, enquêtes *ad hoc*, etc.) la mesure des conséquences sur les attitudes et les comportements de campagnes anti-tabac, ou encore de campagnes en faveur de la tolérance envers les immigrés, peut se révéler très délicate. On retrouve, avec une acuité renforcée, les difficultés précédemment mentionnées :

- mauvaise connaissance du groupe cible avant la campagne (ce qui limite la portée d'analyse avant-après) ;
- possibilité d'effets secondaires non directement perceptibles (report du tabac vers une autre consommation nocive...) ;
- problème de la permanence des effets, etc.²⁷

Enfin, au-delà des politiques de communication sociale au sens strict, l'évaluation de l'impact de politiques publiques « classiques » sur les attitudes et les comportements des Français s'est également développée au cours de ces dernières années. Pour ce faire, l'utilisation de systèmes de suivi sociologique permet de repérer cet impact par rapport aux tendances longues. L'encadré n° 1 présente le système 3 SC²⁸ mis au point et géré par le Cofremca qui permet d'opérer de telles reconnaissances.

IV - 6 La régulation macroéconomique

Le terme de « régulation macroéconomique » renvoie aux diverses politiques publiques ayant pour objet d'influer sur l'évolution des grands agrégats macroéconomiques : produit national, emploi, solde budgétaire, indice des prix, solde commercial, etc. Les instruments utilisés peuvent concerner les politiques publiques précédemment évoquées. Inversement, ces politiques publiques, même lorsqu'elles ne sont pas conduites dans cette perspective, sont susceptibles de peser d'un poids notable sur l'évolution macroéconomique globale. On touche là au problème du caractère mixte des politiques ou, encore, de la multiplicité des objectifs et des effets.

L'évaluation des effets d'une politique de régulation macroéconomique (exemple : relance, dévaluation, blocage des prix) ou des effets macroéconomiques d'une politique spécifique (exemple : aides à l'emploi) a fait l'objet de plusieurs travaux. Ils recourent tous à la même approche : celle d'une simulation *ex post* de la politique, compte tenu des variables d'environnement macroéconomique observées

27. On se reportera, sur ces divers points, à l'article cité de B. Dubois et qui contient en outre, pour le cas des campagnes d'intérêt public, une comparaison des avantages et inconvénients des divers plans de recherche (groupe témoin, avant-après etc.).

28. 3 SC pour Système Cofremca de suivi des courants socio-culturels.

Encadré n° 1

Le système Cofremca de suivi des courants socio-culturels²⁹

* Il s'agit d'un outil sociologique d'observation et de mesure des changements de la société civile française.

* Il prend en compte l'ensemble de la population métropolitaine française de 15 ans et plus.

* Ce qui est observé :

- les sensibilités, les mentalités ;
- les motivations, les systèmes de valeurs ;
- les comportements et les mœurs, dans tout une série de domaines (alimentaire, habitat, loisir, travail, sexualité, etc.).
- les paradigmes, c'est-à-dire les grandes représentations sous-jacentes du monde, de la nature, de l'avenir, des institutions, des formes d'action, partagées à un moment donné par l'ensemble, ou par tel ou tel segment de population ;
- les structures informelles, relationnelles, la position de l'individu par rapport aux structures de son environnement et aux institutions (l'État, l'entreprise, les syndicats, la famille, etc.) ;
- les processus discrets et systèmes d'autorégulation non immédiatement visibles et qui assurent le fonctionnement d'une société.

* Comment :

- par combinaison de différentes approches de recherche sociologique et capitalisation dans le temps, permettant à la fois des analyses diachroniques et synchroniques ;
- analyse socio-historique par analyses secondaires de milliers d'enquêtes — périodes 1950-1970 — et travaux antérieurs de différents sociologues ;
- monographies ethno-sociologiques — de familles, d'entreprises, de groupes sociaux, d'innovations réussies ou échouées, d'événements sociaux, politiques, médiatiques, etc ;
- entretiens psychosociologiques en profondeur sur des segments de population particuliers, des précurseurs, des cohortes, etc ;
- enquête quantitative annuelle reproduite depuis 1974 sur un échantillon renouvelé chaque année de 2 500 Français, représentatif de la population de 15 ans et plus, et alimentant une banque de données socio-culturelles donnant lieu à de multiples entrées et procédures de traitement et structuration permettant de décrire quantitativement les phénomènes de changement et de diffusion des nouvelles valeurs dans la société française.

* Disponibilité et accès à l'information :

- par souscription annuelle donnant droit à l'ensemble des rapports et à l'insertion de questions spécifiques au souscripteur ;

29. Communication au groupe de M. Devez, directeur à la Cofremca.

Encadré n° 1 (suite)

- par séminaires de travail sur l'information socio-culturelle et ses implications dans un domaine donné ;
- élaboration et utilisation de scénarios d'évolution de la société française ;
- par extraction et traitement à la demande d'informations stockées en banque de données.

* Internationalisation

La Cofremca a suscité la création d'observatoires utilisant sa méthodologie dans 18 pays, dont tous les principaux pays européens, regroupés au sein d'une structure commune : RISC. Des programmes de recherche européens, permettant des comparaisons de nature et de rythme d'évolution, ont été lancés depuis 1981 (programme ACE — Anticiper le Changement en Europe).

* Applications principales :

- marketing des produits et services, marketing de la consommation ;
- recherche d'innovation produits/services ;
- marketing socio-politique — Marketing des institutions et services publics ;
- évaluation et pilotage des stratégies, des politiques et des actions ;
- sensibilisation — introduction à l'action systématique sur des ensembles complexes ;
- élaboration des stratégies, des desseins à long terme d'entreprises ou institutions intégrant les dynamiques sociétales ;
- pilotage catalytique, combinant action / évaluation / ajustement.

* Principaux programmes spécifiques dérivés :

- évolution des mœurs alimentaires des Français ;
- travail et changement social ;
- socio-médical ;
- habitat et cadre de vie ;
- international food and drink monitor ;
- médias.

(exemple : conjoncture internationale, niveau du dollar), avec un **modèle macroéconométrique**. Divers exemples de cette approche, portant sur le passé récent, figurent dans l'ouvrage A. Fonteneau et P.-A. Muet (1985), *La gauche face à la crise*, Presses de la Fondation nationale des sciences politiques³⁰. On trouvera dans l'encadré n° 2 un résumé de l'évaluation faite de la relance de 1981-1982.

Cette approche présente plusieurs avantages : facilité de mise en œuvre (dès lors qu'un modèle est disponible, bien sûr !), possibilité de décomposer l'interprétation causale (par exemple, les effets moindres que prévu d'une relance sont dus, pour tel montant, à un accroissement du taux d'épargne des ménages, pour tel autre, à un environnement international plus bas que projeté lors de la formulation de la politique etc.). L'expérience accumulée en la matière conduit à juger la **fiabilité numérique des résultats obtenus convenable, mais sur la seule période de validité des mécanismes des modèles macroéconométriques utilisés, c'est-à-dire 2 ou 3 ans**. L'évaluation avec des modèles macroéconométriques de politiques de « moyen terme » reposant sur des mécanismes « d'offre » plus inertes (exemple : politiques de restructuration industrielle, abaissement du coût relatif capital-travail...) apparaît, en l'état de l'art actuel, beaucoup plus sujet à caution.

30. Plusieurs études analogues ont été publiées dans la revue *Economie et Statistiques*.

LA RELANCE DE 1981-1982 : L'ÉVALUATION DE L'OFCE

L'évaluation de l'OFCE a consisté tout d'abord à répertorier les principales mesures de relance prises (revalorisation du SMIC, mesures sociales, actions en faveur de l'emploi et de l'investissement, mesures fiscales et parafiscales) et à les « traduire » en termes de variables exogènes du modèle OFCE-annuel. Divers types de simulation peuvent être alors effectués :

- avec un environnement international conforme aux prévisions retenues lors de la mise en œuvre de ces mesures ;
- avec l'environnement international observé.

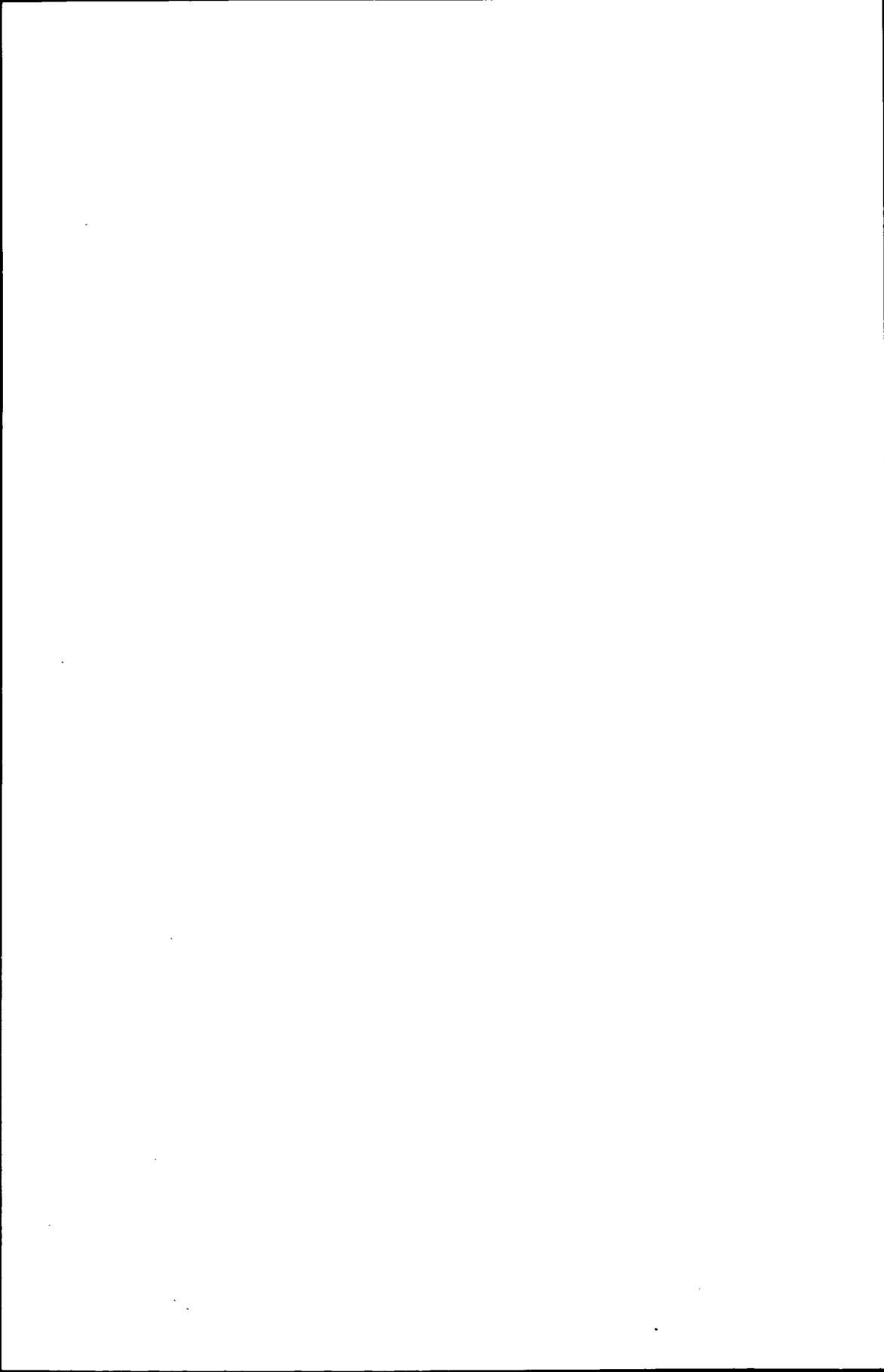
Par comparaison à l'évolution observée, on peut mettre en évidence la contribution des divers facteurs, soit de politique économique, soit d'environnement international, aux résultats atteints. A titre d'illustration on reproduit un tableau mettant en évidence l'impact de la détérioration de l'environnement de l'économie française en 1982 (par rapport à l'environnement prévu).

	Evolution observée* (1)	Impact :		Evolution simulée avec l'environnement prévu en 1982 (4) = (1)-(2)-(3)
		Baisse de la demande et de l'inflation mondiale (2)	Hausse du dollar par rapport à l'écu (3)	
<i>Variations par rapport à 1981 (%)</i>				
PIB (volume)	1,8	- 1,7	- 0,3	3,8
Consommation des ménages (volume)	3,4	- 0,3	- 0,5	4,2
Investissement total (volume)	0,7	- 1,9	- 0,9	3,5
Exportations (volume) ..	- 2,6	- 7,0	+ 0,3	4,1
Importations (volume) ..	6,8	- 4,5	- 0,8	12,1
Prix à la consommation	11,2	- 0,1	+ 2,0	9,3
<i>Niveau</i>				
Chômage (en milliers) ...	1 870	+ 90	+ 10	1 770
Solde des biens et services (MdF)	- 100	- 20	- 24	- 56

* Les volumes sont mesurés aux prix de l'année 1970. Les évolutions observées en 1982 pour l'investissement, les importations et, à un moindre degré, la consommation des ménages diffèrent de celles qui sont mesurées aux prix de l'année précédente. Ces divergences tiennent aux différences de pondération dans les deux modes d'évaluation.

V

**Les enseignements
de quelques expériences nationales**



V

Les enseignements de quelques expériences nationales

V - 1 Evaluation et esprit des peuples

Le mouvement évaluatif n'est pas seulement la mise en œuvre de certaines méthodes de recherche pour apprécier les effets des actions publiques dans tel ou tel domaine, c'est en même temps un ensemble de pratiques administratives et politiques qui définissent son organisation, l'usage qui en est fait et les conséquences qu'il a sur le fonctionnement des institutions. L'observation rapide d'un certain nombre d'expériences nationales, dont la nôtre, montre à quel point l'aspect apparemment universaliste des méthodes de recherche et des techniques statistiques se combine avec le particularisme de chaque système institutionnel et de chaque régime politique. Mis à part les régimes non démocratiques avec lesquels l'évaluation, dont la logique profonde est le doute scientifique, est par nature incompatible, les pratiques d'évaluation expriment l'« esprit des peuples » et leur façon d'utiliser plus ou moins intensément les acquis des sciences sociales comme instrument de démocratie.

Nous commencerons par l'exemple des États-Unis. L'ampleur et la diversité de l'expérience américaine lui confèrent en effet une place particulière. Son étude montre en outre que l'évaluation y est un mouvement fortement institutionnalisé et aux racines anciennes, c'est-à-dire tout le contraire d'une mode. Cela sera confirmé par une brève revue de quelques autres pays dans lesquels l'évaluation s'est implanté selon des voies originales.

On examinera ensuite le cas français en s'efforçant de répondre à deux questions. La première est de savoir pourquoi l'évaluation reste assez peu développée en France. La seconde consiste à s'interroger sur les possibilités d'un développement spontané de l'évaluation dans notre pays.

V - 2 L'exemple des États-Unis

1) La prise de conscience, à la fin des années 1960, de l'échec du Planning Programming Budgeting System (PPBS) est parfois citée comme étant à l'origine du mouvement évaluatif aux USA. En fait, si l'expérience du PPBS a joué un rôle important dans le développement de l'évaluation, celle-ci est issue principalement d'un mouvement bien antérieur et beaucoup plus profond constitué de la recherche en sciences sociales appliquée aux conséquences de l'action gouvernementale¹.

Ce courant, très ancien dans son inspiration, s'est exprimé pour la première fois en rapport direct avec l'action du pouvoir fédéral lors du New Deal. A partir de cette époque, de multiples recherches empiriques vont être menées dans les domaines de l'éducation et des programmes sociaux, bénéficiant des progrès méthodologiques, notamment en matière de mesure, offerts par les études psychologiques menées pour l'armée américaine pendant la guerre. Les années 1950 voient le développement d'évaluations importantes utilisant les méthodes d'enquête et d'analyse statistique mises au point progressivement dans les différentes disciplines des sciences sociales. En effet, ce courant d'évaluation est pluridisciplinaire, il s'enrichit des apports méthodologiques de la sociologie, la science politique, la psychologie, les statistiques, l'économie et le marketing. Il va se trouver considérablement renforcé par l'existence d'un débat permanent dans les années 1960 sur le bien-fondé des réformes lancées sous Kennedy et surtout sous Johnson (la « Great Society ») dans les domaines de l'éducation, la justice, l'aide sociale, l'urbanisme, les transports, le logement et qui vont entraîner une croissance très rapide des dépenses publiques.

Cette décennie est celle de la croissance vertigineuse et du déclin aussi rapide de l'effort de rationalisation connu sous le terme de **Planning Programming Budgeting System (PPBS)**. Ce deuxième courant qui alimente le mouvement évaluatif est inspiré essentiellement par les économistes, renforcés par les spécialistes de recherche opérationnelle. Le PPBS, système intégré d'administration, comporte une phase amont de fixation d'objectifs et de choix des projets reposant sur les techniques d'analyse de système et de calcul économique. Cette phase suppose un travail d'évaluation des effets des programmes publics. Il s'agit là du lien intellectuel le plus direct, même s'il n'est que très partiel, entre l'évaluation de politique et le PPBS. Une différence fondamentale existe néanmoins entre ces deux conceptions. L'évaluation opérée dans le cadre des études analytiques du PPBS est essentiellement une *évaluation a priori des effets attendus* de programmes alternatifs et non une observation et une mesure *a posteriori* des effets réels d'une

1. Pour cette analyse du mouvement évaluatif aux États-Unis nous nous inspirons d'une part de l'étude de E. Chelimsky, *L'évaluation de programmes aux États-Unis*, Politique et Management Public, n° 2, juin 1985, et d'autre part de l'analyse de L. Rouban présentée dans un rapport particulier effectué pour le groupe de travail sur *L'évaluation des politiques publiques aux États-Unis*, ronéo (Centre de recherches administratives et politiques, université de Rennes, décembre 1985). E. Chelimsky décrit notamment la constitution des deux courants à l'origine du mouvement évaluatif. L. Rouban étudie en particulier le lien entre l'évaluation et la vie politique américaine. Nous utilisons également F. Levy, *L'évaluation des politiques publiques aux États-Unis* in EPP. op. cit., ainsi que N. Beckman, ed. *Policy Analysis in Government, alternative to muddling through*, Public Administration Review, 37, n° 3 (1977), p. 221-262.

politique une fois qu'elle a été appliquée (même si le bouclage du système prévoyait *in fine* une comparaison prévisions-réalisations). En outre, l'évaluation que suppose le PPBS repose principalement sur le calcul économique alors que l'évaluation de politique puise plus largement dans différentes sciences sociales.

En 1965, sur la base du succès apparent de la mise en place du PPBS au département de la Défense, Lyndon B. Johnson décide sa généralisation à l'ensemble des départements ministériels et des agences fédérales. Peu de temps après apparaissent à la fois les premiers signes de déconvenue et les premières analyses² d'un échec qui sera consommé en 1971. Parmi les multiples causes de cet échec, certaines ont trait directement aux études analytiques. D'une part, les économistes ont découvert que leur conception normative de la rationalité n'était pas celle qui présidait aux décisions politiques et aux choix budgétaires, d'autre part, il est apparu que leurs ambitions analytiques étaient démesurées compte tenu de l'écart entre les moyens d'évaluation *a priori* dont ils disposaient notamment en matière d'instruments de mesure, de systèmes statistiques et de données disponibles sur les phénomènes en jeu, et ceux qui auraient permis d'effectuer des estimations réalistes des effets attendus des programmes. L'écart entre l'ambition des analystes et la faiblesse de l'information était d'autant plus accentué que les programmes en question correspondaient à des domaines et à des moyens d'actions nouveaux pour l'administration fédérale américaine.

Autrement dit, l'échec du PPBS a fait ressentir aux analystes comme aux autorités l'intérêt de développer l'analyse *a posteriori* et, pour cela de renforcer les systèmes statistiques et les bases de données, et d'une façon générale, la connaissance des processus en œuvre dans les principaux domaines d'intervention de l'État.

2) La lente et précautionneuse démarche de l'évaluation empirique des politiques par le courant de la recherche appliquée, qui s'appuyait sur la connaissance accumulée, l'expérience et la formation progressive des évaluateurs, s'avérait donc finalement plus efficace, y compris comme aide à la décision, que la démarche plus technocratique du PPBS. Mais cette expérience avait apporté, outre la conviction au sein de l'exécutif qu'il fallait poursuivre l'effort d'analyse des politiques, des méthodes telle que l'analyse coûts-avantages utilisables dans l'évaluation *a posteriori*, et aussi des modifications dans les procédures de préparation et d'exécution des budgets à travers des ensembles de programmes hiérarchisés qui constituent depuis autant d'unités d'analyse pour les évaluateurs.

Les années 1970 vont être celles du développement accéléré de l'évaluation et de son institutionnalisation. D'environ 20 millions de dollars à la fin des années 1960, les crédits fédéraux dévolus à l'évaluation sont décuplés 10 ans après. Encore ces chiffres ne représentent-ils qu'une partie de l'effort d'évaluation, puisque celle-ci est aussi le fait des États, des universités, de fondations, de cabinets privés, dont les sources de financement sont diverses.

Parmi les centaines d'évaluations menées annuellement apparaissent des évaluations de grande ampleur et d'une grande exigence méthodologique, notamment

2. Notamment le célèbre article de A. Wildavsky, *Restoring policy analysts from PPBS*, *Public Administration Review*, n° 29, 1969, p. 189-201.

des expérimentations dont l'une des plus célèbres est menée sur le principe de l'impôt négatif (voir partie II), ainsi que de vastes évaluations dans le domaine éducatif ou celui des programmes sociaux.

L'évaluation, et plus globalement l'analyse de politique, se consolident comme discipline à part entière dans le domaine académique. Des programmes doctoraux de formation d'analystes et de chercheurs se constituent dans les écoles et les universités, qui proposent également des programmes de formation continue aux fonctionnaires en activité. L'analyse des politiques devient, en 1975, une des matières obligatoires dans les programmes de maîtrise destinés aux futurs fonctionnaires³.

L'évaluation s'implante dans le système politico-administratif. La branche exécutive se dote de cellules d'évaluation dans les différents départements qui constituent d'importantes banques de données à usage d'évaluation et l'**Office of Management and Budget** devient un grand utilisateur d'évaluations. Il en est de même de la branche législative avec le **Congressional Budget Office** créé en 1974 pour évaluer les projets de lois de finance. Mais c'est surtout le **General Accounting Office (GAO)** qui constitue l'instrument principal d'évaluation auprès du Congrès (voir partie III). Le GAO développe, surtout à partir d'une loi de 1974 qui renforce le statut légal de l'évaluation, une importante activité évaluatrice à côté de sa fonction traditionnelle d'audit financier.

Le mouvement évaluatif, et plus généralement l'analyse des politiques, constitue aux États-Unis, dans les années 1980, un milieu professionnel de plusieurs milliers d'analystes répartis dans les universités, les centres de recherche, les fondations, les cabinets d'études, l'administration fédérale et locale et les services parlementaires. Ces spécialistes se retrouvent au sein de « sociétés savantes »⁴ et s'expriment et publient les résultats de leurs travaux dans plusieurs revues scientifiques spécialisées dont la plus importante est aujourd'hui le *Journal of Policy Analysis and Management*. Les résultats des études sont fréquemment repris dans la grande presse, et le débat politique lui-même est alimenté par les résultats des études. L'ampleur du mouvement et son institutionnalisation ne doivent pourtant pas masquer que celui-ci n'est à l'abri ni d'une certaine banalisation ni de certains aléas dus aux changements politiques et aux querelles d'écoles sur la nature et l'utilisation de l'évaluation.

3) Sous les présidences démocrates de Kennedy et de Johnson, s'engager à évaluer les programmes sociaux était un moyen de donner des garanties à certaines parties de l'opinion et du Congrès. D'autre part, l'administration qui ne vivait pas en situation de rationnement budgétaire voyait dans l'évaluation le moyen d'améliorer, au plan technique, ses modalités d'intervention. La préoccupation d'efficacité l'emportait donc sur celle d'efficience. Cette dernière va, au contraire, se renforcer sous l'administration Nixon. Celle-ci qui avait, d'une part, des éléments d'appréciation des résultats de certains programmes de la période

3. *Guidelines and standards for professional Master Degree program in public affairs-public administration*, National Association of Schools of Public Affairs/Public Administration, 1974.

4. Policy Studies Organization (science politique), Society for the Study of Social Problems (sociologie), Society for the Psychological Study of Social Issues (psychologie), Evaluation Research Society (pluridisciplinaire).

précédente, d'autre part, le souci de ne pas se laisser entraîner dans une vague supplémentaire de programmes publics, utilisa l'évaluation pour redéfinir son intervention davantage sous forme d'incitation que de prestations directes.

L'administration Nixon qui a lancé de grandes évaluations de nature expérimentale fut accusée de l'avoir fait également parce que c'est une forme d'évaluation qui permet d'attendre, éventuellement longtemps, ses résultats et peut donc servir d'alibi scientifique à l'attentisme.

La période de méfiance vis-à-vis de l'exécutif qui suit le départ de Nixon voit le **renforcement des pouvoirs évaluatifs du Congrès**. L'arrivée de Reagan, marquée par une volonté de rupture avec la croissance de l'intervention de l'État, se traduit à la fois par un coup d'arrêt donné à certains programmes sociaux et un ralentissement de l'effort d'évaluation mené par l'exécutif. Ce ralentissement va être en fait modéré et de courte durée. D'une part, parce que l'évaluation est pluraliste. Ainsi, un programme d'aide aux adolescents ayant quitté le domicile familial qui avait été lancé par Carter et qui était très critiqué par l'Administration Reagan fut-il évalué par le GAO à la demande du Congrès⁵. Ce programme ayant fait l'objet d'une évaluation très favorable, le Congrès, loin de le supprimer, augmenta son budget de 50 %. D'autre part, l'exécutif lui-même prit assez vite conscience du rôle que pouvait jouer l'évaluation pour réaliser ses projets de déréglementation, de réduction des programmes d'intervention et d'amélioration de la productivité dans les services publics. Ainsi, les crédits relatifs aux sciences sociales de la National Sciences Foundation, qui avaient été supprimés par l'administration Reagan dans un premier temps, ont été rétablis deux années plus tard. Et un effort considérable est fait pour accentuer la pratique de l'évaluation au niveau subfédéral (États et grandes municipalités).

Alors qu'en matière de système de gestion, on avait vu le PPBS de Kennedy/Johnson, puis le « Management By Objective » de Nixon et le « Zero Base Budgeting » de Carter, le développement sur moyenne période du mouvement évaluatif ne semble pas profondément affecté par l'alternance politique. Les changements d'orientation politique du pouvoir modifient les usages de l'évaluation et influence les types de démarche évaluative qui vont prévaloir, notamment du côté de l'exécutif. Toutefois, l'évaluation s'est déployée et installée au sein du système politico-administratif américain par l'action de conditions favorables qui tiennent à ce système lui-même mais aussi à des causes plus larges qui en font une composante de la démocratie américaine.

4) L'administration fédérale est constitutionnellement soumise à la double autorité de la Présidence et du Congrès. L'institutionnalisation de l'évaluation tant du côté du législatif que de l'exécutif peut largement s'interpréter comme un élément de la « course aux armements » entre les deux pouvoirs, cette compétition étant tempérée par l'existence d'instances intermédiaires comme le GAO dont les travaux d'évaluation contribuent au dialogue entre les deux branches du gouvernement.

5. U.S. General Accounting Office, Federally supported centers provide needed services for runaways and homeless youths, GAO/IPE — 83-7, 26 septembre, 1983, cité par E. Chelimsky, op. cit.

L'évaluation joue également un rôle d'intégration et de régulation dans un système administratif dont la cohésion n'est pas assurée, comme elle l'est en France, par de multiples dispositifs d'intégration tel qu'un droit administratif particulier, la planification nationale, le système des grands corps ou les inspections ministérielles.

Enfin, les valeurs culturelles propres à la société américaine font une grande place à l'empirisme et à l'approche inductive de la décision. De plus, l'utilisation de la science pour éclairer l'action publique est une valeur largement partagée et ce, depuis les origines de la démocratie américaine. Les sciences sociales jouissent d'un statut élevé notamment depuis la démonstration de leur apports pratiques lors de la Seconde Guerre mondiale. Elles vont donc pouvoir « entrer dans le siècle », dialoguer entre les différents pôles de pouvoir et servir de base au développement de l'évaluation.

V - 3 Autres expériences étrangères : R.F.A, Pays-Bas, Royaume-Uni, Suède, Suisse

Si les États-Unis ont incontestablement joué un rôle d'avant-garde dans le développement de l'évaluation, plusieurs expériences étrangères, parfois plus modestes, fournissent également des enseignements intéressants. On passera successivement en revue le cas de la République fédérale d'Allemagne, des Pays-Bas, du Royaume-Uni, de la Suède et de la Suisse.

1) **L'administration fédérale allemande**⁶ a connu, comme la plupart des pays occidentaux, un mouvement de rationalisation incluant la conception de mécanismes de planification et de contrôle dans la fin des années 1960. Ce mouvement a coïncidé avec la formation en 1965 d'une coalition entre les chrétiens démocrates et les socio-démocrates et l'acceptation par les premiers, sous l'impulsion des seconds, d'une plus forte intervention de l'État dans les activités économiques et sociales. Ce changement de point de vue est considéré comme l'origine, avec un décalage de temps, des activités d'évaluation qui se sont développées à partir de 1969.

6. Des compléments sur le cas de la R.F.A. peuvent être trouvés dans les documents suivants :

H.U. Derlien, *Program evaluation as an approach to the management of public policies in the Federal Republic of Germany*, Symposium on management of public policies, ENA, Paris, 27 octobre 1980.

H.U. Derlien, *Implementations probleme : Bürokratische ineffizienz oder politische Programm fehler ?* dans *Verwaltungsarchiv*, 1984.

H. König, *L'évaluation de la politique en République fédérale d'Allemagne*, conférence du Groupe européen d'administration publique, 3-6 septembre 1985, Louvain.

G. Färber, *Evaluation and financial control in the Federal Republic of Germany*, conférence du Groupe européen d'administration publique, 3-6 septembre 1985, Louvain.

Dès le début des années 1970, on constate dans plusieurs ministères la création d'unités de recherche socio-économique qui, dès leur origine, se consacrent à l'évaluation *a priori* et *a posteriori*. La plupart de ces évaluations sont accomplies sous forme de recherches sur contrat, mais ces unités spécialisées discutent les plans de recherche, fournissent des données, passent les contrats et suivent le travail des équipes de recherche. La création de ces unités a été encouragée par la réforme budgétaire de 1969 qui permettait au ministère des Finances d'exiger une étude économique des ministères présentant des programmes nouveaux. De telles unités furent créées au ministère de l'Agriculture, des Transports, de l'Aide au Développement, de la Technologie. Cette dernière s'est surtout consacrée à l'évaluation des institutions de recherche et des programmes de recherche financés sur fonds publics. Une organisation de l'évaluation des politiques de communication a également été mise en place à l'Agence fédérale sur les relations avec la presse et à l'Agence fédérale pour l'information de santé.

Le Parlement a joué un rôle important dans la mise en place de dispositifs d'évaluation. Ainsi qu'il a été constaté dans le cas des États-Unis, les doutes d'une large partie de l'opinion, reflétés par les médias, sur le bien-fondé de nombreux programmes d'intervention étatique sont à l'origine d'une exigence croissante du Parlement quant à l'évaluation des effets de ces programmes. Depuis 1970, de très nombreux votes de lois et de programmes dans les domaines économiques, sociaux et éducatifs ont été assortis d'une obligation pour le gouvernement fédéral d'organiser une évaluation de la mise en œuvre et des effets obtenus.

Des expérimentations ont simultanément été conduites dans les domaines de l'éducation et de la santé. Le choix de ces domaines n'est pas étranger à des raisons politiques. En effet, l'éducation et la santé sont constitutionnellement attribuées aux régions et ne relèvent pas du gouvernement fédéral qui n'est autorisé dans ces domaines qu'à définir un cadre général. L'expérimentation a constitué pour le gouvernement fédéral un moyen de lancer des réformes sur des territoires hors de sa compétence, en passant des accords administratifs entre ses services et ceux des gouvernements locaux. La justification d'une telle procédure étant de multiplier les expériences afin de clarifier les bases d'une réglementation générale qui, elle, est du ressort du gouvernement fédéral.

La répartition des tâches en matière d'évaluation des effets des politiques est très nette quant à la réalisation des études, plus complexe quant à leur prescription et leur utilisation. La réalisation des travaux d'évaluation est considérée comme une tâche spécifique qui ne peut être accomplie, sauf dans les cas simples, par les institutions traditionnelles de contrôle ou par les services de l'administration. Lorsque certains ministères disposent d'unités spécialisées dans l'évaluation, ces unités, en général, ne conduisent pas elles-mêmes les recherches. L'évaluation est fondamentalement externe sous forme de recherche sur contrat avec des centres de recherche universitaires ou des sociétés d'études.

L'initiative de l'évaluation est généralement le fait du Parlement. Elle peut aussi venir de la Cour des comptes ou de la Chancellerie mais rarement du ministère chargé de la politique. La fonction des unités spécialisées dans les ministères est plutôt d'administrer l'évaluation externe et de faire un rapport au Parlement sur la base des études d'évaluation en proposant des modifications à la politique évaluée.

La fonction d'évaluation s'inscrit donc dans un dispositif institutionnel qui reste peu centralisé. Les services de la Chancellerie se limitant à une fonction de coordination, et le ministère des Finances n'ayant pas comme en France un rôle intégrateur, chaque ministre a, surtout dans les gouvernements de coalition, une grande autonomie, y compris dans ses rapports avec le Parlement.

Le fait que le Parlement ait joué un grand rôle dans ce processus peut expliquer que l'évaluation n'est pas très populaire dans l'administration fédérale où elle est souvent présentée sous la même terminologie que les contrôles traditionnels. Elle est ainsi parfois associée dans les esprits aux inspections ayant pour objet le respect des règles et la mesure des performances individuelles. Dans ces conditions, la résistance à l'évaluation peut exister et ne sera surmontée que s'il existe une base légale à sa réalisation ou si les critiques portées contre la politique sont très fortes. Dans ces cas où la probabilité d'une évaluation négative est élevée, on observe que les problèmes méthodologiques s'accroissent pour l'évaluateur, par défaut de coopération dans la fourniture des données et l'ouverture des dossiers. La tentation est forte ensuite pour les services de contester la validité de telles évaluations surtout si celles-ci sont globales et négatives. Les évaluations comportant une analyse approfondie de la mise en œuvre de la politique et des facteurs explicatifs des effets mesurés paraissent mieux acceptées par les administrations.

Enfin, la tendance semble être actuellement en faveur d'évaluations très sélectives, au besoin en recourant à des études préalables de faisabilité, utilisant des techniques robustes et non plus sophistiquées, faisant une large part à l'analyse de la mise en œuvre de la politique.

2) Les années 1970 ont vu un développement assez considérable des travaux de type PPBS aux Pays-Bas⁷. Toutefois, ce sont surtout les nouvelles politiques qui ont été examinées, conformément à l'attrait que suscitait à l'époque les tâches de conception. La troisième étape de la boucle évaluation→formulation→mise en œuvre était donc assez largement négligée.

Le souci de juger des actions publiques effectivement menées s'est installé progressivement à la fin des années 1970 et au début des années 1980.

En réponse à des critiques visant certaines actions ou bien parce que l'obligation de réaliser une évaluation était prévue légalement, des travaux ont été lancés dans le domaine de l'environnement, sur la réforme universitaire, sur la mise en œuvre par les administrations locales des politiques décidées centralement (entre autres).

7. Cette analyse s'appuie sur les documents suivants :

— E. Sorber, *Some experiences with program evaluation in the Dutch central government*, Workshop on Policy Evaluation, Louvain, septembre 1985.

— Th. Van der Tak, *Policy evaluation*, Workshop on Policy evaluation, Louvain, septembre 1985.

— H.P. Potman, *Evaluation of the Noise Nuisance Act in the Netherlands : the use of Acceptance of policy in evaluation research*, Workshop on Policy Evaluation, Louvain, septembre 1985.

— M. Horrevoets et F. Vonks, *Communication*, Workshop on Policy Evaluation, Louvain, septembre 1985.

— J. Van Geest et A. Ringeling, *Administrative Impact Assessments*, Workshop on Policy Evaluation, Louvain, septembre 1985.

M. Sorber a par ailleurs présenté une communication au groupe de travail.

Jusqu'à une période récente, ces démarches étaient restées assez partielles. Elles étaient souvent l'œuvre de chercheurs universitaires dont les préoccupations ne répondaient pas toujours à celles des décideurs. Une certaine tradition s'était ainsi dégagée mais sans qu'aucune organisation spécifique fut créée.

La situation a évolué récemment avec la création en 1982 au sein de la direction du Budget d'une cellule de coordination méthodologique et de diffusion de travaux en matière d'analyse des politiques publiques. Une nouvelle phase de la recherche est ouverte. Après des évaluations spécifiques employant des techniques lourdes, fortement orientées vers la quantification, deux axes de réflexion se dessinent : l'un met l'accent sur la mesure des effets, l'autre se concentre sur l'analyse des organisations publiques et des procédures d'action. Par ailleurs, les analystes essaient de délivrer des messages plus directement utilisables par les décideurs ; quant aux services gouvernementaux, leur approche de l'évaluation se modifie en raison :

- de la tendance à la privatisation et à la déréglementation ;
- de clauses prévoyant pour certaines lois soit une durée limitée (sunset), soit une évaluation périodique ;
- d'une attitude plus tournée vers la production de résultats.

Ces services tentent de développer leur propre activité évaluative, le développement créé à la direction du Budget ayant un rôle de soutien et de diffusion des méthodes et des travaux. Sont ainsi menées des actions de formation, diffusés des ouvrages de méthodes et un périodique : *Belei Analyse* (Analyse de Politiques).

Un aspect de l'évaluation aux Pays-Bas est l'échange qui est organisé entre les experts du gouvernement central, les autorités locales, les chercheurs universitaires et les sociétés d'études privées. Pour promouvoir cette idée, le ministre des Finances a invité à une conférence sur l'état de l'art en matière d'analyse de politique des décideurs et des analystes représentant les quatre groupes précités. Cette réunion a donné lieu à la création d'un groupe permanent et d'un programme de développement et d'animation du mouvement évaluatif.

3) A certains égards, le contexte dans lequel s'est développée l'évaluation au Royaume-Uni⁸ présente des similitudes avec le cas français. Du côté de la demande d'évaluation il y a, à la différence des États-Unis, une tradition de « dédain culturel » de la part des gouvernants et des hauts fonctionnaires pour les sciences sociales et leur contribution potentielle à la décision publique. D'autre part, contrairement à bien des idées reçues, les politiques gouvernementales sont préparées dans un cercle décisionnel très centralisé, constitué de hauts fonctionnaires des administrations centrales et des groupes d'intérêt qui sont leurs partenaires permanents. Le développement de l'évaluation trouve ses racines dans le désenchantement qui résulte des espoirs excessifs mis, dans les années 1960-1970, dans la rationalisation des choix par l'usage du calcul économique et des systèmes budgétaires intégrés. L'idée d'évaluation a également pénétré dans les collectivités territoriales et le secteur des organisations autonomes de santé, mais ces secteurs ont des moyens limités à y consacrer.

8. L'exposé de l'expérience britannique est principalement issu de la présentation de M.J. Spackman, *L'évaluation au Royaume-Uni*, avec un commentaire de J.C. Thoening, in EPP, chap. 3, ainsi que de B.W. Hogwood L.A. Gunn, *Pc. Analysis for the real world*, Oxford University Press, 1984.

Du côté de l'offre, la tradition britannique est celle de sciences sociales relativement cloisonnées et centrées sur les problèmes théoriques. L'intérêt pour l'application des sciences sociales aux problèmes pratiques d'administration est récent et apparaît à une partie de la communauté scientifique comme le moyen de développer une approche de la recherche et de l'enseignement plus appliquée et plus interdisciplinaire. L'analyse de politique en tant que discipline a mis de nombreuses années à être reconnue dans les milieux scientifiques et à être enseignée dans quelques universités, instituts polytechniques et écoles de gestion. Elle fait également l'objet aujourd'hui de programmes de formation pour fonctionnaires dans ces institutions et au « Civil Service College ». Dans le cas anglais, il est commode de classer les procédures d'évaluation selon leur objectif.

a) **Les procédures d'évaluation comme moyen de promouvoir ou de faciliter la mise en œuvre d'une politique nouvelle**

Il n'existe pas actuellement de procédure centralisée prévoyant l'évaluation de politique. Il n'en n'a pas toujours été ainsi. Le « **Program Analysis and Review** » (PAR) a fonctionné pendant plusieurs années, jusqu'à une date récente. Chaque année, les ministères devaient définir avec le Trésor quelles étaient les politiques publiques qu'il convenait de réexaminer. Pour chaque politique étudiée, une commission était constituée sous l'égide du ministère concerné avec la participation du Trésor et des autres ministères intéressés. La procédure était secrète : les PAR n'étaient pas annoncés officiellement et les rapports n'étaient pas publics. La procédure des PAR s'est avérée très utile. Il y a été mis fin quand les ministères ont été dotés de capacités d'évaluer ou de faire évaluer eux-mêmes leurs politiques. Les capacités d'évaluation sont très variables d'un ministère à l'autre. Certains disposent de services spécialisés qui participent à l'évaluation. La plupart des ministères supervisent des évaluations faites par les universités ou les institutions extérieures.

Il n'existe plus de service central chargé de faire des évaluations pour contribuer à l'élaboration de nouvelles politiques. De 1973 à 1983, un tel organisme, le « **Central Policy Review Staff** », faisait partie du secrétariat général du gouvernement et était composé pour moitié de fonctionnaires détachés pour deux ou trois ans et pour moitié de personnalités scientifiques venant de diverses origines. La disparition de cet organisme unique a laissé place à des divisions plus petites, proches du Premier ministre, qui étudient l'efficacité des ministères et leur fonctionnement et pas seulement les grands problèmes de politiques nationales. Une tradition particulière au Royaume-Uni est constituée par les Commissions spéciales et les Comités d'enquêtes qui mêlent fonctionnaires et personnalités compétentes extérieures à l'administration.

Outre les travaux effectués ou financés par l'administration, une grande partie de l'évaluation visant à promouvoir de nouvelles politiques est bien entendu réalisée en toute indépendance par des instituts de recherche et des groupes de pression.

b) **Les procédures d'évaluation dans le cadre d'un réexamen périodique**

Ce type de procédure n'est pas systématique mais certaines mesures sont assorties d'une décision de réexamen au terme d'une période déterminée ou de façon

récurrente. Il ne s'agit pas seulement d'un réexamen lié à la procédure budgétaire (bien que celle-ci puisse conduire à des études plus approfondies) mais d'une évaluation spécifique à échéance fixe. Il s'agit par exemple de la décision de privatiser les procédures de règlement de contentieux d'assurance ou de la politique de fixation de la marge acceptable dans les marchés publics de défense nationale n'ayant pas fait l'objet d'un appel d'offres. Dans ce dernier cas, le réexamen du taux de marge et son mode de calcul est effectué tous les deux ans par une commission indépendante.

c) Les procédures d'évaluation comme moyen d'améliorer les processus de décision et de gestion des dépenses

Les évaluations de ce type sont, si elles portent sur des objectifs limités, effectuées par le ministère chargé de la politique, ou, si elles portent sur des objectifs plus larges, supervisées ou réalisées par un organisme distinct de celui qui est responsable de la politique.

Comme exemple du premier type, on peut citer les évaluations (assez classiques) réalisées dans le domaine des transports (étude d'axes nouveaux ou aménagés, de l'effet des péages ou de nouvelles voies de circulation sur l'activité industrielle...). Des réexamens plus complets sont effectués par des services de recherche internes à l'administration mais externes au ministère décideur et par des centres de recherche universitaires ou des sociétés d'études ou d'audits privés. **L'appel à ces experts extérieurs s'est beaucoup renforcé dans les dernières années.** Ces études portent sur des domaines très variés tels que l'évaluation des établissements scolaires, la politique pénale, l'efficacité des organisations de santé ou les aides à l'industrie.

L'audit externe de l'administration a évolué d'autre part en direction de l'évaluation. L'audit externe de l'administration centrale est effectué par l'Exchequer and Audit Department (E & AD) sous l'autorité du Comptroller and Auditor General, qui est l'homologue du Premier président de la Cour des comptes. Traditionnellement consacrés à la vérification des comptes et à la dénonciation des irrégularités et des gaspillages manifestes, ces contrôles sont étendus depuis le National Audit Act de 1983 à des missions d'études « portant sur les économies réalisées, l'efficiences et l'efficacité ». L'Exchequer and Audit Department a vu ses moyens renforcés quantitativement et qualitativement notamment par l'apport d'experts extérieurs. Les rapports de l'E & AD sont examinés par la Commission des comptes publics de la Chambre des communes qui généralement les publie ainsi que ses propres recommandations après avoir, le cas échéant, entendu publiquement les hauts fonctionnaires concernés par ces rapports.

Une telle procédure et la publicité qui l'accompagne donnent beaucoup de poids aux évaluations. Il est intéressant de signaler que la Commission des comptes publics regroupe des membres de tous les partis composant le Parlement et qu'elle est normalement présidée par un ancien ministre des Finances appartenant à un parti de l'opposition.

d) Les procédures d'évaluation comme moyen de faire connaître les résultats obtenus à l'opinion publique

Outre la Commission des comptes publics, d'autres commissions parlementaires suivent l'action de chaque département ministériel. Elles peuvent étudier spécifiquement telle ou telle politique et entendre les représentants de ministères concernés. Le Livre blanc sur les dépenses publiques publié chaque année est utilisé de plus en plus comme outil d'évaluation et comporte de plus en plus de mesures de résultats.

Au total, l'expérience britannique se caractérise par un apprentissage de la démarche évaluative qui l'a fait passer d'une activité centrale et globalisante à une activité polycentrique et liée directement à l'action concrète. L'évolution récente s'est traduite par l'élimination des risques de monopole et par l'ouverture du système d'évaluation à de multiples instances : auprès du Premier ministre, dans les ministères, à l'E & AD, dans les centres de recherche universitaires et dans les sociétés d'études et d'audit privées. D'une certaine façon l'évaluation est en voie de se banaliser. Ce n'est plus une activité nouvelle dotée de vertus magiques mais un acte de bonne gestion parmi d'autres.

e) La Suède⁹ est l'un des pays à économie de marché dans lequel la part des dépenses publiques a le plus augmenté dans les trente dernières années. C'est aussi l'un de ceux dans lesquels la remise en cause de cette croissance s'est manifestée le plus tardivement.

Homogène, de petite dimension, urbanisée et industrialisée, la société suédoise a manifesté jusqu'à une période récente un consensus fort quant au développement du secteur public, essentiellement dans les domaines sociaux et de l'éducation. Ce secteur public est largement décentralisé puisque les deux tiers des budgets publics relèvent des comtés et des municipalités. Dans les années 1960 ont été mis en place des systèmes de planification et de contrôle de type PPBS qui ont largement été utilisés comme moyens de décentralisation de l'administration. L'application de ces méthodes n'a pas freiné l'accroissement des budgets publics mais a contribué à l'efficacité de la gestion, peut-être parce que, à la différence de nombreux autres pays, ces méthodes n'ont pas été appliquées à de vastes administrations centrales mais à des organismes autonomes et de taille restreinte, et avec un accent mis sur les systèmes d'information et non sur les études analytiques qui ont connu les mêmes difficultés de réalisation que dans les autres pays.

Ce consensus sur le développement des politiques publiques, notamment dans le domaine social, et le sentiment partagé d'une assez bonne gestion des organismes publics, peuvent expliquer la relative lenteur du développement des travaux d'évaluation en Suède. D'autres facteurs peuvent également être évoqués. Le premier est que les petits pays sont relativement désavantagés. Bien que le nombre de questions sur lesquelles une expertise est utile à la décision soit équivalent à celui d'un grand pays, les ressources disponibles en matière de recherches économiques et sociales dans un pays comme la Suède sont beaucoup plus limitées.

En second lieu, le caractère très décentralisé de la décision publique constitue une forme de substitut à l'évaluation. Les citoyens étant proches des décideurs, la

9. Voir : — B. Ch. Ysander, *L'évaluation des politiques publiques en Suède*, in EPP, chap. 4.
— R. Fremfors, *Social research and governmental commissions in Sweden*, American Behavioural Scientist, vol. 26, n° 5, 1983.

pression exercée sur le gouvernement central pour qu'il évalue les politiques a été relativement moins forte que dans d'autres pays, compte tenu également du courant redonnant davantage de décentralisation et un allègement des réglementations nationales. Bon nombre des études d'évaluation effectuées au cours de la période récente ont en effet été centrées sur des modifications de la structure des pouvoirs, à côté des études portant sur la réglementation du marché du travail, les aides sociales, les traitements médicaux, les systèmes de retraite ou l'éducation qui constituent avec les études sur les politiques de l'énergie, les gros bataillons des travaux suédois.

L'organisation des travaux d'évaluation explique en effet la lenteur des progrès de l'évaluation de l'efficacité des politiques et surtout leur caractère peu apparent.

Le principal véhicule de l'évaluation des politiques en Suède est le système des **commissions nationales**. Les lois importantes et les modifications majeures de politiques sont généralement élaborées par une commission nationale nommée par le gouvernement, composée de parlementaires issus des grands partis, de représentants des groupes sociaux (syndicats, fédérations patronales), de fonctionnaires qui assurent le secrétariat et d'experts extérieurs.

Étant donné qu'il y a toujours deux ou trois cents commissions qui se réunissent simultanément, et ce en moyenne pendant trois ans, la quasi-totalité des experts du pays apportent leur contribution aux travaux des commissions. La tâche des experts porte essentiellement sur l'évaluation des politiques passées et sur l'analyse des nouvelles options. Mais, d'une part, les directives du gouvernement limitent la portée des évaluations réalisées et, d'autre part, la principale fonction de ces commissions est de préparer la voie à un consensus parlementaire. Les travaux des experts peuvent être ainsi utilisés plus comme des armes politiques que comme une base de décision objective et avoir une faible visibilité en tant qu'information indépendante.

Une autre source importante d'évaluation est constituée par les **Commissions temporaires de recherche et de développement** créées par les principaux ministères mais jouissant d'un statut semi-indépendant et composées de représentants du ministère et des agences nationales concernées ainsi que d'experts extérieurs. Les ministères peuvent également faire appel directement à une prestation extérieure pour une évaluation, procédure actuellement en développement.

La troisième source d'évaluation est le **Bureau national de contrôle** dont la vocation est de contrôler l'efficacité des agences nationales qui sont les instruments d'action des ministères, ceux-ci étant réduits à un état-major de 40 à 100 personnes. Le Bureau national de contrôle a évolué progressivement d'un contrôle externe de l'efficacité de chaque agence, celles-ci ayant par ailleurs leurs propres systèmes de contrôle interne, vers une évaluation directe des politiques en débordant du cadre d'activité de chaque agence.

Enfin, le **niveau local** qui pèse d'un poids considérable, mais dont l'activité est pour une part réglementée au niveau national, est très surveillé au plan de l'efficacité de la gestion. Il ne fait pas l'objet d'évaluations de politiques très fréquentes sauf dans les grandes municipalités. De tels travaux sont encouragés par les deux fédérations nationales des collectivités locales qui disposent de services de recherche qui leur sont propres.

Cette organisation très fragmentée et l'affaiblissement du rôle du Parlement ne sont pas favorables à un développement rapide de l'évaluation ni à des évaluations approfondies et de grande ampleur. Seul un examen plus détaillé permettrait de comprendre si cet état de choses tient à la faiblesse de la demande sociale pour l'évaluation ou au fait que cette demande est au moins partiellement satisfaite par un substitut constitué d'une part par l'usage systématique du contrôle de gestion dans les agences nationales et collectivités territoriales, d'autre part, et surtout, par une forte participation des experts ou représentants de l'opinion dans le cadre des commissions nationales et des organisations décentralisées.

f) En Suisse¹⁰, l'évaluation des politiques publiques à l'aide des instruments de la recherche scientifique a commencé durant les années 1970. Au début, l'initiative d'évaluer était prise par le Gouvernement, le Parlement, l'administration ou encore par les chercheurs indépendants. Le mouvement a été notablement renforcé grâce aux programmes nationaux de recherche de la **Fondation nationale suisse**. Cette institution de recherche subventionnée par le gouvernement a fait réaliser un nombre élevé d'études d'évaluation dans différents domaines. Un accent particulier a été mis sur les effets régionaux de mesures budgétaires et de programmes de régulation ainsi que sur l'analyse des procédures de décision.

Ces travaux n'ont pas encore répondu complètement aux espoirs qu'ils avaient suscités en ce qui concerne leur traduction politique. Ceci tient en partie au fait que de nombreux programmes de recherche sont encore en cours de réalisation.

Pour s'en tenir aux études menées et terminées durant la période 1975-1983, en dehors des programmes de recherche, on peut citer notamment :

- la réduction de la vitesse limitée dans les agglomérations de 60 à 50 km/h (cf. partie II) ;
- la ceinture de sécurité ;
- les effets des autoroutes sur les régions concernées ;
- l'équité du financement des routes soit par la fiscalité, soit par la tarification.

Les fonctions des évaluations sont multiples : production d'informations nouvelles, préparation de décisions à venir, renforcement du consensus, légitimisation de la politique suivie. Les méthodes employées sont diverses, analyse coûts-avantages, analyse d'efficacité, expérimentation, simulation, construction d'indicateurs d'impact qualitatifs et quantitatifs.

Les évaluations sont menées tant par l'administration fédérale, cantonale,... elle-même que par des experts extérieurs et conduisent à des échanges fructueux.

10. Dr. Armin Jans, *Some Notes on Policy Evaluation in Switzerland*, Workshop on Policy Evaluation, Louvain, septembre 1985.

V - 4 Le cas de la France

1) La recherche appliquée et les techniques d'aide à la décision sont, en France également, à l'origine de l'évaluation mais dans des proportions et dans un contexte très différents.

La recherche dans les sciences sociales s'est développée en France dans un rapport distant avec le pouvoir. La recherche économique elle-même ne s'est implantée dans les circuits de décision publics qu'à partir des années 1950¹¹. Ce lien entre la recherche empirique et la décision politique et administrative reste jusqu'aux années 1960 épisodique et propre à quelques secteurs privilégiés : contributions de psychologues et sociologues dans le domaine de l'éducation, du Centre de sociologie des organisations en matière de réformes administratives par exemple. Des travaux que l'on pourrait qualifier aujourd'hui d'évaluations apparaissent mais ne sont pas élaborés ni utilisés dans cette perspective.

En parallèle, le courant du calcul économique¹² et de l'aide à la décision se développe dans les années 1960 sous l'impulsion des ingénieurs-économistes de l'administration, notamment dans les secteurs de la Défense, de l'Équipement, et de l'Énergie. Ces secteurs font alors l'objet d'une forte demande et il y a un relatif consensus sur les efforts d'investissement qu'ils nécessitent.

Le lancement de la rationalisation des choix budgétaires, en 1968, marque l'extension de l'analyse économique *a priori* dans de nombreux domaines en même temps que la multiplication des services d'études économiques et de programmation.

Cette nouvelle phase se déroule toujours en période de croissance mais l'ampleur et la diversification de la demande des services publics ne permet plus d'obtenir un consensus quant aux projets à réaliser. Les divergences sur les estimations des effets attendus des politiques étudiées commencent à s'accroître, conduisant à l'utilisation de méthodes multicritères qui sont davantage utilisées comme schémas de négociation que comme instruments de calcul.

A partir de 1975, les conséquences de la crise vont se faire sentir sur les modalités même de prise de décisions. Les études *a priori* s'avèrent de plus en plus insérées dans les processus de négociation entre organisations et groupes sociaux. Elles se diversifient dans leur nature et certaines incluent des évaluations *ex post* de politiques¹³. L'éclatement du mouvement RCB, comme celui du PPBS aux États-Unis, conduit à une certaine prise de conscience de l'insuffisance des données et de la connaissance des processus par lesquels les programmes publics produisent leurs effets dans le public. Là aussi un certain rapprochement se réalise entre les deux courants avec le développement limité, mais incontestable, dans les années 1975/1985, de travaux d'évaluation *ex post* menés principalement à l'initiative de

11. F. Fourquet, *Les comptes de la puissance*, Encres, 1980.

12. R. Poinard et B. Walliser, *Évaluation et évolution des études d'aide à la décision*, Bulletin RCB, n° 50, septembre 1982.

13. R. Poinard et B. Walliser, *Rapport d'évaluation des études prioritaires*, Bulletin RCB, n° 43, décembre 1980.

l'administration mais auxquels contribue beaucoup plus largement que par le passé la communauté scientifique en économie, sociologie, science politique et gestion¹⁴. En 1986, l'évaluation est en France une activité qui progresse faiblement et n'est pratiquement pas organisée ni institutionnalisée au sein du système politico-administratif. La production qu'on peut estimer grossièrement à quelques dizaines d'études par an, avec d'assez grandes différences entre acteurs, est à une échelle encore artisanale tant au sein des administrations qui en sont les premiers prescripteurs et réalisateurs que dans les centres de recherche. Les évaluations réalisées sont peu utilisées par les mass médias et les hommes politiques, elles ne constituent pas un élément majeur de préparation des décisions publiques. Un certain rapprochement entre les deux courants a été réalisé du côté de la recherche universitaire par le progrès du travail empirique et une moindre acuité des luttes entre écoles de pensée (néo-classiques contre marxistes en économie, structuro-fonctionnalistes contre marxistes en sociologie, comportementalistes contre freudiens en psychologie, etc.). D'autre part, certaines parties de l'administration, notamment les comités de recherche du Commissariat général du plan, ont joué un rôle pilote pour la mise en rapport des services d'études de l'administration et des centres de recherche *extra-muros*.

Une troisième composante qui contribue à l'originalité du mouvement évaluatif en France est constituée par le rôle des corps de contrôle (voir partie I). La préoccupation de l'évaluation s'est fait jour à la Cour des comptes et dans les inspections générales où elle représente une activité encore réduite mais significative. Malgré ces limitations, ce courant d'évaluation joue un rôle important par son ambivalence. D'une part les contrôles ayant un objectif d'évaluation contribuent au développement de celle-ci, mais d'autre part l'existence même et le mode de fonctionnement des corps de contrôle peuvent constituer un frein à l'institutionnalisation d'une évaluation de type scientifique. En effet, non seulement de larges parties de l'opinion des milieux administratifs et politiques ne perçoivent pas clairement la différence entre l'évaluation et les contrôles traditionnels, mais nombreux sont ceux qui considèrent plus ou moins implicitement que si une fonction d'évaluation doit exister, elle doit rester interne à l'administration, ce qui conduit plus ou moins à assimiler évaluation et modernisation des contrôles administratifs.

III apparaît donc que l'évaluation en France existe de façon significative depuis le début des années 1980 mais n'est pas encore reconnue au plan institutionnel comme une activité normale, régulière et organisée. Elle reste une activité sporadique même si de nombreux signes manifestent qu'elle est l'objet d'attentes diverses, éventuellement excessives. Cet état de fait s'explique par le jeu d'un certain nombre de freins spécifiques qui limitent son développement en France.

2) A cet égard, les difficultés méthodologiques mises en avant pour expliquer le développement limité de l'évaluation en France doivent être appréciées à leur juste valeur. Pour l'essentiel, les obstacles de type méthodologique naissent de contradictions opposant les conditions de mise en œuvre des politiques publiques et les exigences de la recherche scientifique. Sous ce dernier aspect, évaluer une

14. Sur le développement et la nature des études d'évaluation en France, voir J.-P. Nioche et R. Poinard, *L'évaluation des politiques publiques, en France*, in EPP, op. cit.

politique revient à établir une causalité entre les effets (attendus ou non) et la cause que constitue la politique. C'est certainement la procédure expérimentale qui offre les meilleures garanties pour établir la relation causale dans de bonnes conditions. Malheureusement, il est vrai que les plans de recherche de ce type supposent satisfaire de nombreuses conditions (cf. II ci-dessus).

Toutefois, les évaluations menées depuis des années dans plusieurs pays montrent que les sciences sociales offrent de nombreuses solutions techniques qui, pour peu que l'on ne tombe pas dans le fétichisme méthodologique, permettent d'apporter des éléments de réponse sinon dans tous les cas et de façon totalement fiable, du moins de manière satisfaisante et avec un degré de clarté très supérieur à celui du débat politique usuel.

Les difficultés méthodologiques n'expliquent pas vraiment le faible développement de l'évaluation en France. Celui-ci est plutôt le fruit de l'environnement administratif, politique et culturel qui limite le recours à l'évaluation des politiques publiques et rend difficile l'utilisation de certaines méthodes.

3) **Les obstacles administratifs paraissent importants mais peuvent être surmontés par certaines réformes.** Le premier de ces obstacles est l'absence de gestion par programme. Pourtant des tentatives ont eu lieu en ce sens au cours des vingt dernières années : les « programmes prioritaires » du plan, les « budgets de programmes » mis en place dans le cadre budgétaire au niveau de chaque département ministériel, les « plans d'action » élaborés par certains services administratifs. Sauf en ce qui concerne les programmes prioritaires, l'insertion de ces instruments dans les procédures de négociation et leur impact sur les décisions sont restés limités¹⁵. S'ils ont pu contribuer d'une part à rendre plus habituel le raisonnement en terme de moyens et de résultats de l'action administrative et, d'autre part, à produire une information nouvelle, les outils de ce type n'ont pas empêché que la préparation des budgets, les arbitrages gouvernementaux, les débats parlementaires, la gestion des services, les contrôles, s'exercent toujours selon la logique du budget de moyens. Le degré de précision des politiques, de leurs objectifs, des moyens qui leur sont attribués, de leur enjeu, est donc faible en moyenne, ce qui complique la tâche des évaluateurs. Il s'agit là d'une des grandes différences entre la France et les États-Unis où la gestion par programme est la règle.

Par ailleurs, le fonctionnement des structures administratives réalise un équilibre « bas » en matière d'évaluation. La rareté des évaluations et la faiblesse de leur impact décisionnel ne conduisent pas à la création d'équipes spécialisées. En conséquence, le partage des tâches d'évaluation entre divers services opérationnels ou d'études, inspections et corps de contrôle, qui pourrait être une source de richesse, a du mal à dépasser le stade de l'éparpillement et à opérer des synergies effectives. Le même processus est observable au plan financier puisque les cas de politiques prévoyant les moyens de leur évaluation dès leur lancement sont rarissimes. Or, l'évaluation sans moyens ne peut montrer qu'elle peut être efficace et constituer, par là, un facteur d'économie.

15. P. Poinard, *Les budgets de programmes, quinze ans après*, Economie et Prévision, n° 71, 1985/5.

D'une façon générale, l'administration française souffre sans doute de certaines lacunes dans son système de formation. A l'exception de l'économie (et de manière non systématique) les futurs hauts fonctionnaires sont peu mis en contact avec la recherche contemporaine dans les différentes sciences sociales. La formation permanente ne venant pas compenser les handicaps de la formation initiale, les méthodes de recherche en sciences sociales et leurs développements actuels, donc les ressources qu'elles peuvent constituer dans l'action administrative, sont en définitive peu connues.

Enfin, l'évaluation peut apparaître à certains comme redondante avec les multiples formes de contrôles administratifs, politiques et juridictionnels dont font l'objet leurs actions.

4) **les obstacles politiques et culturels** sont par nature plus insaisissables mais jouent sans doute un rôle majeur dans les réticences que rencontre l'évaluation. Peu développée dans l'administration, l'évaluation n'a pas trouvé de conditions favorables à son implantation dans d'autres sphères, notamment à cause de la prééminence de l'exécutif sur le législatif. Il s'agit là d'une différence fondamentale avec les États-Unis où la compétition entre exécutif et législatif a conduit à une sophistication croissante des mécanismes d'évaluation de part et d'autre.

En France, la situation privilégiée de l'administration en matière d'informations économiques et sociales ainsi que la relative faiblesse des mass medias dans leur rôle d'évaluateur de l'action des pouvoirs publics ont freiné le développement de l'évaluation.

De façon structurelle, la nature unitaire et très centralisée de l'État ainsi que l'existence d'un droit administratif très cohérent sont interprétés comme limitant les possibilités de modulation des politiques en fonction de conditions locales (au sens large). Les occasions d'évaluation dans des conditions d'expérimentation « naturelle » par comparaison entre différentes options de politiques sont donc réduites. A cet égard, la décentralisation peut constituer une chance pour l'évaluation et notamment l'expérimentation de politiques¹⁶.

Plus profondément encore, certaines représentations mentales, qui conditionnent les comportements administratifs et politiques, sont des freins puissants à un développement de l'évaluation.

Notre culture intellectuelle conduit à donner l'avantage à la déduction sur l'induction, au jugement normatif sur la recherche empirique, aux tâches de conception sur les fonctions opérationnelles. L'évaluation, qui repose sur l'observation, le doute et la mesure sur le terrain se heurte à ces valeurs et à ces traditions. Elle s'est heurtée à **une représentation dominante, bien que déclinante, de l'État comme promoteur infallible de l'intérêt général**. Cette idéologie consiste à regarder ou proclamer une politique comme satisfaisante, c'est-à-dire conforme à l'intérêt public, du moment qu'elle a été décidée régulièrement par une autorité légitime, ce qui exclut implicitement tout besoin d'évaluation.

16. A condition qu'il n'y ait pas une course à l'appropriation de l'évaluation.

Naturellement, et pour le même type de raisons, les méthodes d'évaluation les plus satisfaisantes scientifiquement se heurtent dans notre pays aux obstacles les plus nombreux. Ainsi, l'expérimentation est peu répandue, non seulement à cause des difficultés techniques qu'elle présente, qui ne sont pas propres à la France, mais parce qu'elle oblige ceux qui veulent y recourir à admettre publiquement qu'ils peuvent avoir des doutes sur l'efficacité des mesures qu'ils proposent, et à affronter le risque d'être accusé d'injustice, d'immoralité ou d'illégalité pour non respect du principe d'égalité de tous devant la loi.

V - 5 Les limites d'un développement spontané de l'évaluation dans notre pays

L'évaluation est donc, en France, une activité encore endémique et peu insérée dans les processus de décision. On a dit qu'elle faisait l'objet de nombreuses attentes dont certaines reposent sur des conceptions floues, hétérogènes et parfois contradictoires. Le développement modeste mais incontestable de l'évaluation dans la période récente, malgré l'ampleur des obstacles administratifs, politiques et sociologiques laisse toutefois supposer que des mécanismes plus profonds sont en jeu, comme le montrent certains exemples étrangers.

La question fondamentale qui se pose est de savoir si ces mécanismes sont susceptibles d'engendrer, spontanément et sans intervention volontariste de la puissance publique, un développement continu et solide de l'évaluation des politiques publiques. Pour prendre une image économique, peut-on compter sur le « libre jeu de l'offre et de la demande » pour assurer le développement de l'évaluation, sa qualité et sa bonne insertion dans les débats et les choix politiques¹⁷ ? En quelque sorte, l'évaluation a-t-elle un marché ?

1) L'« offre » d'évaluation est diversifiée mais cette abondance apparente masque une mauvaise distribution des capacités techniques et organisationnelles nécessaires à une évaluation de qualité. Il n'existe en France aucune institution ayant pour vocation principale de faire de l'évaluation de politique. Il y a très peu d'organismes dans lesquels l'évaluation atteint une masse critique et une permanence suffisantes pour accumuler l'expérience et faire progresser l'expertise. Le pouvoir législatif ne dispose pas des moyens requis pour faire des évaluations de qualité. Il est en effet caractérisé en France par la faiblesse de ses capacités d'expertise propres¹⁸. Les services de l'Assemblée nationale sont dotés de moyens d'études encore limités dont les efforts actuels de modernisation sont orientés

17. Sur cette analyse du « marché » de l'évaluation, voir J.-P. Nioche et R. Poinard, *L'évaluation des politiques publiques en France*, in EPP, op cit.

18. J.-L. Quermonne et L. Rouban, *L'évaluation et le contrôle politique*, in EPP, op. cit.

prioritairement vers la collecte et l'utilisation des informations produites par l'administration. Le Parlement se trouve donc dans une situation de dépendance de fait vis-à-vis de l'administration quant aux données sur les effets des politiques gouvernementales. La création de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (loi du 8 juillet 1983) a introduit une innovation notable, mais celui-ci a pour objet l'évaluation *a priori* et non *a posteriori*.

Le Conseil économique et social se trouve dans une situation analogue malgré l'apparition notable d'évaluations dans ses travaux récents.

L'évaluation à la Cour des comptes est le résultat des programmes de travail que se donne cette juridiction. La part qu'elle fait à l'évaluation est encore limitée compte tenu du poids de ses métiers traditionnels.

Les inspections générales ministérielles¹⁹ se trouvent dans une position similaire bien qu'il existe des différences importantes entre elles. Les contrôles débordant sur l'évaluation de politique existent dans presque toutes les inspections mais avec des fréquences variables, qui sont assez élevées, tout en étant minoritaires, à l'Inspection générale des finances, ou à l'Inspection générale de l'industrie.

D'une façon générale, les évaluations réalisées dans les corps de contrôle restent d'ampleur limitée et leurs méthodes sont fondamentalement marquées par la tradition de vérification de régularité et d'audit financier.

Lorsque des méthodes nouvelles, telles que les enquêtes, sont utilisées, leurs modalités restent assez éloignées des normes professionnelles des milieux scientifiques, notamment compte tenu des contraintes de délais.

Ces normes ont davantage pénétré les services d'études des ministères. Mais, là aussi, l'évaluation constitue une activité marginale par rapport aux études de surveillance de l'environnement et d'éclairage des choix. Elle semble relativement plus implantée dans les secteurs correspondant à des interventions de l'État plus récentes (environnement, culture) ou plus conflictuelles (affaires sociales) et dans les services d'analyse à vocation interministérielle (Direction de la prévision, Commissariat général du plan).

Les services opérationnels effectuent eux-mêmes un certain nombre d'évaluations qui semblent plus souvent limitées à l'analyse de leur propre fonctionnement et des productions qui en découlent.

L'orientation est généralement opposée dans les évaluations réalisées par des évaluateurs extérieurs à l'administration. La presse ou les sociétés d'études mettent généralement l'accent sur la collecte d'informations par enquête dans le public sur les effets des politiques.

Enfin les centres de recherche (universités, grandes écoles, centres indépendants) contribuent de manière significative à la production d'évaluations mais, pour chacun d'eux, il s'agit le plus souvent d'une activité occasionnelle ou secondaire. Elle est, pour les laboratoires spécialisés dans une discipline, l'occasion de tester une théorie et, pour les centres spécialisés dans un secteur, une forme, parmi d'autres, de connaissance du terrain.

19. Voir P. Milloz, op. cit.

Non seulement l'offre d'évaluation est le fait d'évaluateurs pour lesquels elle constitue une activité marginale, mais leurs capacités d'évaluation sont contraintes par leur activité principale. Ainsi les points forts des corps de contrôle sont leur crédibilité auprès des décideurs, leur disponibilité, le caractère opérationnel de leurs recommandations et leur facilité d'accès aux informations internes à l'administration. De manière opposée, les centres de recherche de type universitaire extérieurs à l'administration disposeront en général d'une plus grande indépendance²⁰, seront plus aptes à opérer des remises en cause fondamentales, à collecter l'information externe à l'administration et à mettre en œuvre des méthodes de type scientifique.

2) **La demande d'évaluation** qui en principe pourrait être aussi diverse que l'offre est en fait très concentrée : l'administration y joue un rôle majeur alors que le Parlement, la presse ou les publics concernés sont rarement à l'origine d'évaluations. Bien qu'il soit parfois difficile de départager, dans les évaluations lancées par l'exécutif, celles qui l'ont été par l'administration et celles qui l'ont été par les responsables politiques, il semble que les services opérationnels soient les premiers prescripteurs. Si l'administration joue un rôle prépondérant dans la demande d'évaluation, celle-ci s'exprime de façon peu organisée. Il existe de vastes zones d'ombre, à côté de secteurs faisant l'objet d'évaluations multiples. Ces demandes multiples n'apparaissent pas coordonnées à des fins de vérification ou de complémentarité mais plutôt comme le reflet du souci d'acteurs désireux d'avoir leur information propre sur un sujet ou de répondre à une évaluation par une autre. L'essentiel de la demande se situant au niveau de ministères ou de parties de ministères, il n'est pas étonnant que les politiques interministérielles soient peu évaluées. L'organisation de la Cour des comptes par chambres sectorielles ne l'incitait pas à faire exception à cette tendance. Un décret du 11 février 1985 permet à la Cour de conduire des travaux relevant de plusieurs chambres et facilitera donc l'étude de politiques inter-organismes.

Les centres de recherche ont eux aussi un rôle de prescripteur pour les évaluations qu'ils font à leur propre initiative, mais dans le cadre de moyens généralement limités. **D'une façon générale, les acteurs sociaux autres que l'administration participent faiblement à la demande réelle d'évaluation.** Cela est probablement dû à la faiblesse de leurs moyens financiers et aux difficultés qu'ils rencontrent pour accéder à l'information.

3) Les mécanismes d'ajustement entre l'offre et la demande ne favorisent ni l'insertion de l'évaluation dans les processus de décision ni l'amélioration de la qualité des études ni une contribution croissante au débat public sur les politiques.

Les pratiques actuelles d'évaluation sont caractérisées par l'importance de l'auto-saisine et de l'auto-évaluation. L'auto-saisine correspond aux cas dans lesquels un évaluateur définit lui-même le champ et les objectifs d'une évaluation. L'auto-saisine n'est pas en soi un obstacle à la qualité mais peut fréquemment correspondre à des situations dans lesquelles soit il n'y a pas de demande, soit il n'y a pas coïncidence entre les termes de l'offre et ceux de la demande potentielle, donc une faible chance de voir l'évaluation servir à la décision. L'auto-évaluation,

20. Mises à part les contraintes liées aux activités sous contrat.

c'est-à-dire les cas d'évaluation menée à leur initiative, par les services opérationnels des politiques qu'ils mettent en œuvre, est un cas relativement fréquent. Il s'agit là de travaux qui peuvent servir à la gestion des services et à une responsabilisation de leur personnel mais qui comportent un risque de biais important qui a déjà été souligné.

L'importance de ces phénomènes relativise l'usage que l'on peut faire de la notion de marché pour décrire le mouvement évaluatif en France. Il y a une offre et une demande mais la plupart des évaluations ne font pas l'objet d'un échange entre un demandeur d'information et un offreur d'expertise. En outre, le débat public correspondant aux politiques évaluées bénéficie assez peu de ces travaux. D'une part parce que l'administration, qui joue un rôle prépondérant dans la production d'évaluation, ne publie qu'une part très faible des travaux correspondants. D'autre part parce que les évaluateurs qui publient leurs travaux, tels que les universitaires, le font souvent dans des termes impropres à un débat ouvert aux non-spécialistes.

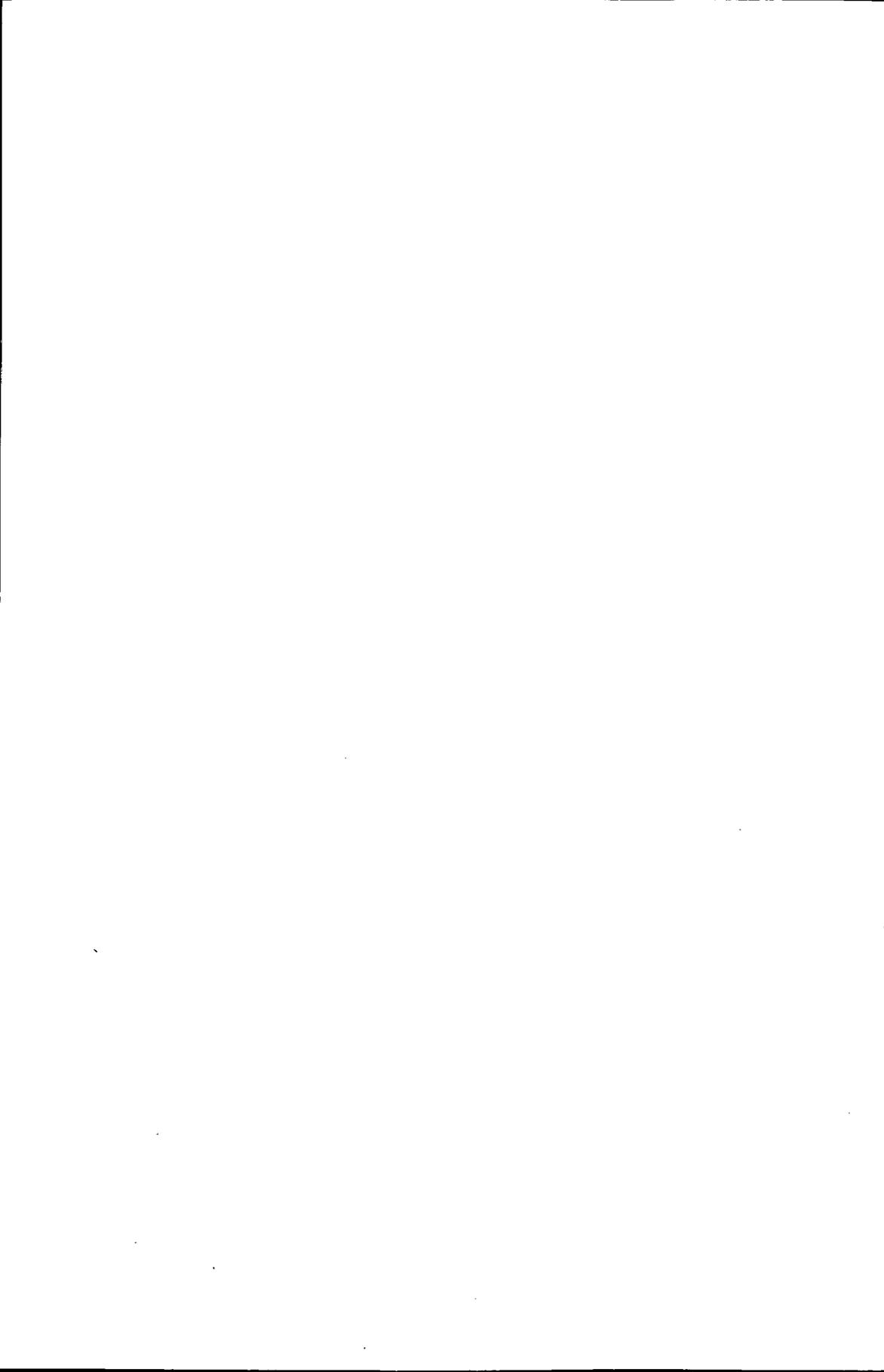
4) L'évaluation n'est ni une activité réglementée ni un véritable marché. Aucune règle ne rend l'évaluation obligatoire²¹, ne définit la manière de choisir quelles politiques doivent être évaluées en priorité, ne suggère des méthodes ou ne désigne des évaluateurs potentiels. A vrai dire, l'évaluation n'est pas une catégorie : les objectifs poursuivis, la nature des rôles, les méthodes mises en œuvre, l'utilisation des résultats ne sont pas codifiés par la règle ni par les usages.

Inversement, l'évaluation n'est pas un véritable marché. L'offre est peu structurée et insuffisante, la demande est excessivement concentrée, les échanges sont faibles et fragmentés. Il ne s'agit d'ailleurs pas d'une activité marchande par essence, même si elle ne peut se développer sans modalités de financement appropriées. On ne peut guère compter sur un élargissement naturel de la demande solvable pour « tirer » le mouvement évaluatif. Il est peu réaliste d'attendre un accroissement des flux d'évaluation d'une multiplication de demandes exprimées cas par cas par les hommes politiques. Ceux-ci sont plus tournés vers l'avenir que vers l'examen du passé, leur horizon temporel est souvent plus bref que celui exigé par l'évaluateur, etc. Mais les responsables politiques peuvent trouver un intérêt à des mesures plus générales permettant de mettre en place un dispositif d'évaluation assurant un flux régulier d'évaluations et mobilisable à la demande sur des projets spécifiques. Les médias sont potentiellement intéressés par l'évaluation mais assez rarement au point de financer eux-mêmes sa réalisation. Les informations relatives aux résultats des politiques gouvernementales sont exploitées le plus souvent de façon ponctuelle sous forme de « scoop moral » comme pour l'affaire des avions renifleurs. Il est probablement possible d'amener les médias à passer d'une recherche stricte d'un spectaculaire sans lendemain, comme l'exploitation annuelle du rapport de la Cour des comptes, à une surveillance permanente de l'action gouvernementale. Mais là aussi il est peu réaliste d'attendre cette transformation d'une évolution naturelle des médias. De la même manière, il ne faut pas espérer un accroissement spontané important de la demande d'évaluation par les administrations et, de toute façon, ces demandes resteront limitées dans leurs

21. Mis à part quelques cas de lois prévoyant explicitement qu'elles devraient faire l'objet d'une évaluation (ex : la loi sur le prix unique pour le livre).

objectifs : aucune administration ne réalisera ou ne fera réaliser une évaluation susceptible de compromettre son « fond de commerce », c'est-à-dire de remettre en cause radicalement la ou les politiques qui constituent son portefeuille d'activités et justifient les moyens dont elle dispose.

Les forces en présence ainsi que les logiques d'action des parties prenantes ne permettent donc pas d'obtenir de manière spontanée un développement rapide et satisfaisant de l'évaluation. Par le mot satisfaisant, nous voulons dire deux choses. D'une part que le respect de normes méthodologiques et déontologiques garantisse la qualité et la crédibilité des évaluations auprès des décideurs et du public afin de conduire à leur prise en compte dans les choix publics. D'autre part que le dispositif global d'évaluation soit « ouvert », c'est-à-dire qu'il assure aux différentes forces sociales la possibilité de faire ou de faire faire des évaluations ; aux évaluateurs qualifiés qu'il permette de répondre à ces demandes ; enfin que les citoyens et leurs représentants aient accès aux évaluations réalisées. Au stade actuel, le développement de l'évaluation relève donc pour partie d'une intervention volontariste de la puissance publique. Une telle orientation ne signifie ni une prise en charge par l'État de l'activité d'évaluation ni une réglementation détaillée, mais suppose la mise en œuvre d'un nombre relativement restreint de mesures visant à faciliter l'émergence d'un dispositif pluraliste de production et d'utilisation de l'évaluation.



VI

Développer l'évaluation des politiques publiques en France Cinq propositions



VI

Développer l'évaluation des politiques publiques en France

Cinq propositions

Le développement de l'évaluation des politiques dans notre pays apparaît hautement souhaitable, à la fois pour affermir la maîtrise et l'efficacité des actions de la puissance publique et pour éclairer le dialogue social.

Il est susceptible, en particulier, de provoquer ou d'accompagner de manière utile l'adaptation de ces politiques, de leurs domaines ou de leurs modalités, à l'évolution de l'environnement économique et social.

Le mouvement des esprits et des structures est propice mais les chances d'un développement spontané, immédiat, de l'évaluation, hors de toute intervention volontariste de la puissance publique, paraissent limitées. La dispersion et l'hétérogénéité des compétences comme des intérêts rendent improbable l'enclenchement automatique d'une dynamique cohérente. Ce constat ne signifie pas pour autant qu'il faille viser une prise en charge colbertiste de l'activité d'évaluation. Il implique, par contre, que soient engagées quelques actions résolues, propres à créer les conditions structurelles nécessaires pour un développement dans lequel l'État ne serait plus, à terme, qu'un acteur parmi d'autres.

L'analyse faite par le groupe, conformément au mandat qui lui avait été fixé, conduit à formuler cinq propositions principales :

- rendre les politiques publiques évaluables ;
- organiser les compétences ;
- renforcer l'évaluation des politiques au sein de l'administration ;
- accroître la capacité d'évaluation externe à l'administration et favoriser sa diffusion dans la société ;
- identifier les thèmes prioritaires.

Pour chacune de ces propositions, un certain nombre d'orientations spécifiques sont définies. Leur mise en œuvre détaillée appelle des dispositions et décisions complémentaires, notamment institutionnelles, qu'il n'appartenait pas au groupe de travail de fixer. Un souci est présent dans l'ensemble de ces propositions : celui d'aboutir à un développement pluraliste de l'évaluation, assis sur la production de travaux de qualité.

VI - 1 Rendre les politiques publiques évaluables

Il a été souligné à plusieurs reprises que les difficultés propres à l'évaluation sont très loin d'être uniquement techniques. Des obstacles plus fondamentaux se rencontrent à l'amont, pour la reconnaissance même de « l'objet » à évaluer. Si certaines conditions minimales ne sont pas vérifiées au stade de la formulation de la politique, l'activité d'évaluation peut être très malaisée voire impossible à conduire. Par rapport à la situation française actuelle, des progrès doivent être recherchés dans cinq directions complémentaires.

1) Il convient tout d'abord de **formuler aussi systématiquement que possible les politiques de manière à en favoriser l'évaluation**. Les pratiques « sédimentaires », par développement à la marge d'un système, l'absence d'objectifs précis et quantifiés, le caractère implicite des hypothèses d'environnement, peuvent engendrer de graves difficultés pour la conduite et l'évaluation des politiques concernées. Il serait donc fort utile, *a contrario*, de rechercher une organisation des politiques par programme, une spécification des objectifs opérationnels, un affichage des hypothèses d'environnement.

2) Il existe une **forte complémentarité entre les études d'aide à la décision, préalables à la formulation et à l'exécution de la politique, et d'évaluation des résultats ex post**. Ces études sont loin d'être systématiques. L'absence d'une connaissance satisfaisante des caractéristiques des groupes concernés par la politique, des mécanismes de mise en œuvre, des effets attendus, est fortement préjudiciable à la validité d'une évaluation *a posteriori*. L'argument selon lequel l'étude *a priori* ralentit l'action ne doit pas être surestimé. Les cas où les politiques doivent être appliquées en véritable urgence pour des raisons techniques sont en fait minoritaires.

3) **L'évaluation se déroulera de manière d'autant plus efficace que son organisation est prévue très tôt, dès la formulation de la politique**. Cette observation concerne, tout d'abord, le dispositif de collecte des informations nécessaires à l'évaluation, soit par recours à des sources existantes (comptabilité, systèmes statistiques), soit par création d'instruments *ad hoc* (enquêtes). Dans ce domaine l'improvisation *ex post*, si elle peut être nécessaire, est souvent coûteuse d'un point de vue financier, mais aussi pour la validité des résultats. Il convient par ailleurs de prévoir à l'origine les moyens de l'évaluation, financiers et humains. Cette pratique est courante aux États-Unis pour tout nouveau programme.

4) L'évaluation d'une politique publique doit faire l'objet d'un **engagement**. Ceci implique que l'on définisse des **échéances d'évaluation**, appelant éventuellement une **procédure d'extinction de la politique**. Une telle discipline limiterait les risques liés à une pratique excessivement discrétionnaire de l'évaluation. Elle constituerait un frein à l'empilement, souvent dénoncé dans ce rapport, de mesures prises successivement par les pouvoirs publics sur un même domaine et dont la cohérence et l'efficacité d'ensemble peuvent être mises en doute.

5) Le recours à l'**expérimentation** constitue une voie privilégiée pour l'évaluation des politiques publiques, notamment les politiques incitatives et réglementaires. Sa pratique peut, dans certaines situations, soulever des difficultés juridiques liées

notamment à la notion d'égalité devant le service public. Il conviendrait de reconnaître précisément ces difficultés et les cas où elles sont réellement insurmontables. En outre, la décentralisation des pouvoirs peut être un facteur favorable au développement de l'expérimentation, en facilitant la mise en place, dans le cadre réglementaire et législatif existant, de dispositifs différenciés suivant les régions. Le cas échéant, la réalisation d'expérimentations pourrait justifier un aménagement *ad hoc* de ce cadre.

6) De ce point de vue, l'évaluation des politiques publiques doit prendre explicitement en compte le fait de la **décentralisation** et le partage des responsabilités entre pouvoirs centraux et locaux qui en résulte. Toutefois, si la décentralisation peut être une chance pour l'évaluation en multipliant les cas de figure, elle comporte le risque d'une segmentation, avec un développement de sources d'information et d'outils méthodologiques locaux inutilement spécifiques, ce qui réduirait la comparaison et la généralisation des résultats obtenus. **Une réflexion sur l'évaluation des politiques menées par les autorités décentralisées est donc nécessaire.** Une première recommandation peut être formulée, qui consiste à rechercher un accord entre les différents acteurs, nationaux et locaux, sur les méthodes, tout en favorisant le pluralisme des initiatives.

VI - 2 Organiser les compétences

La partie proprement technique de l'évaluation appelle des compétences spécifiques. Des efforts sont nécessaires pour les développer en France. Ils doivent prendre en compte l'existant et éviter dispersion et double emploi.

1) Le développement ou la mise en place des **formations** nécessaires, initiales ou permanentes, devrait en priorité **s'appuyer sur les organismes compétents existants**, sans nécessairement chercher à créer des structures *ad hoc* dont la pérennité ne serait plus assurée en régime de croisière. Il n'est pas dans l'objet du rapport de dresser une liste exhaustive de ces points d'appui dont on se bornera à mentionner quelques uns, tels que l'École nationale de la statistique et de l'administration économique, le Centre d'enseignement de la statistique appliquée à la médecine et à la biologie générale, l'Institut de management public, le Centre HEC-ISA. C'est par leur intermédiaire que devraient se diffuser les techniques insuffisamment utilisées en France, en matière d'expérimentation, d'analyse statistique ou de modélisation microéconométrique.

2) La constitution de **pôles d'étude de dimension critique, susceptibles de capitaliser compétences et expériences**, notamment à l'extérieur de l'administration, est une priorité. Elle suppose une mobilisation des instruments existants à la disposition des pouvoirs publics (financement de la recherche, contrats, appels d'offres) en évitant toute dispersion. On ne peut que déplorer le manque de rigueur

méthodologique qui caractérise souvent les appels d'offres de recherche, par exemple en matière de politiques sociales, et qui conduisent à des programmes fourre-tout. De même, l'attribution des moyens financiers participe trop souvent de préoccupations distributives. Le développement de l'évaluation ne requiert pas nécessairement de dégager des moyens nouveaux, si l'on veut bien mobiliser de manière efficace l'existant. Dans le même ordre d'idées, la puissance publique devrait inciter les centres de recherche publics et parapublics, ainsi évidemment que ses propres services d'études et de statistiques, à rééquilibrer leurs activités en faveur des tâches d'évaluation.

3) Enfin, les procédures de collaboration entre prescripteurs administratifs et évaluateurs, externes ou internes, doivent être améliorées. Trop souvent, les objectifs poursuivis par le prescripteur administratif sont insuffisamment précisés. De même, les conditions d'échange d'informations sont mal définies et une faible attention est portée à la lisibilité des travaux. Il en résulte une insatisfaction mutuelle et la tentation pour chaque partie de camper sur son propre territoire, celui de l'opérationnel ou celui de la connaissance. L'établissement de « cahiers des charges » précis, entre prescripteurs et évaluateurs, permettant notamment un suivi attentif des travaux, permettrait de pallier ces difficultés.

VI — 3 Renforcer l'évaluation des politiques publiques au sein de l'administration

Un développement de l'évaluation purement externe à l'administration aurait toute chance de n'avoir qu'un impact assez faible sur l'action de la puissance publique, directement responsable de la conception et de l'exécution des politiques et détentrice d'avantages d'information importants. Un progrès équilibré de l'évaluation dans notre pays, porteur de changements effectifs, suppose donc qu'un effort particulier soit fait au sein de l'administration elle-même. Il doit s'appuyer sur plusieurs principes.

1) Le premier concerne l'initiative de l'évaluation des politiques publiques réalisée par l'administration. La prise en compte du principe de subordination hiérarchique des services administratifs à l'égard de l'autorité ministérielle quant à l'initiative de décision, conduit naturellement à affirmer également la primauté de celle-ci en matière de prescription d'évaluation et à rejeter l'idée d'une seule auto-évaluation administrative par auto-saisine des services eux-mêmes. Pour éviter toutefois des pratiques excessivement discrétionnaires, il conviendrait de privilégier des procédures d'engagement, conduisant à une réalisation irréversible une fois le principe de l'évaluation retenu, et d'améliorer les conditions de diffusion à l'intérieur comme à l'extérieur de l'administration.

2) Les évaluations réalisées au sein de l'administration doivent se conformer aux deux règles déontologiques de **reproductibilité** des protocoles et de **séparation** entre les responsabilités d'évaluation et d'exécution.

La première règle a pour objet de rendre les évaluations vérifiables, en garantissant des conditions suffisantes de transparence. Ces conditions supposent un affichage des protocoles suivis comme une maîtrise des techniques utilisées. Pour atteindre ce dernier objectif, il sera le plus souvent nécessaire d'associer diverses compétences, par appel aux spécialistes concernés de l'administration et à des experts extérieurs, étant entendu que les institutions exerçant actuellement des responsabilités de contrôle (Cour des comptes, Inspection générale des finances, Inspections ministérielles) devraient jouer un rôle particulièrement moteur dans cette dynamique.

La règle de séparation constitue également un impératif, même si son application prend en compte les conditions particulières propres à chaque évaluation comme cela a été souligné dans le rapport. Une attitude *a priori* trop favorable à l'auto-évaluation est susceptible de conduire à une indépendance insuffisante du responsable de l'évaluation par rapport aux services opérationnels chargés de la mise en œuvre de la politique. Il faut réaffirmer qu'une fois la mission confiée à l'évaluateur, celui-ci doit être indépendant dans le choix de ses méthodes et la formulation de son jugement sur les résultats de la politique.

3) Rechercher des règles systématiques sur la diffusion des résultats de l'évaluation serait illusoire. On peut par contre recommander des **schémas gradués** de diffusion assurant au minimum, en cas d'obligation de confidentialité sur l'information de synthèse ou les conclusions, une possibilité d'accès à l'information de base collectée, en s'inspirant de règles suivies en matière d'information statistique ou de déontologie des sondages.

4) Enfin, il paraît important d'assurer un **caractère interministériel** au développement de l'évaluation au sein de l'administration. Les modalités retenues à cette fin peuvent être variables. **Le groupe a reconnu l'intérêt d'un rattachement particulier du développement de l'évaluation au sein de l'administration auprès du Premier ministre, éventuellement par le biais d'une instance spécifique.** Quelle que soit la solution institutionnelle retenue, placer l'évaluation à un niveau interministériel est une réponse, en termes techniques et de légitimité, au caractère souvent global des politiques et de leurs résultats. Le dépassement des seuls champs ministériels est en outre indispensable pour :

- améliorer la qualité des travaux en soumettant ceux-ci à une plus large confrontation et expertise ;
- favoriser la diffusion de leurs résultats au sein de l'administration et éviter la rétention d'informations ;
- permettre une capitalisation des expériences ;
- favoriser l'intégration des résultats des évaluations dans les procédures d'arbitrage interministérielles, notamment de caractère budgétaire.

VI — 4 Accroître la capacité d'évaluation externe à l'administration et favoriser sa diffusion dans la société

Le développement de l'évaluation au sein de l'administration est un point de passage obligé pour assurer sa diffusion dans la société. Les modalités évoquées précédemment participent donc de ce dernier objectif. Elles doivent être explicitement finalisées en ce sens et complétées par des dispositions spécifiques.

1) **Le renforcement des pôles de compétences externes à l'administration comme une publicité accrue des évaluations faites au sein de l'administration sont évidemment des moyens importants pour fournir des outils et des méthodes aux évaluateurs externes.** Un parallèle peut être fait avec la politique menée depuis le rapport Lenoir en matière de diffusion de l'expertise économique. La constitution volontariste de centres extérieurs à l'administration (OFCE, IPECODE) a effectivement permis un transfert de technologie et autorisé une remise en cause du monopole administratif traditionnel dans ce domaine. Des voies similaires peuvent être suivies en matière d'évaluation, à cette différence près que la capacité d'expertise de l'administration en matière d'évaluation se situe probablement à l'heure actuelle en deçà de celle-ci, qui existait en matière d'analyse macroéconomique lorsque ces centres externes ont été constitués.

2) **Le potentiel d'évaluation paraît actuellement très déséquilibré au préjudice du législatif, contrairement à la situation d'autres pays.** Trois leviers sont susceptibles d'être utilisés pour corriger cette carence :

— développer les moyens propres à la disposition de l'Assemblée nationale et du Sénat en termes de personnels, de mode de traitement de l'information, de crédits, ce qui permettrait en particulier l'appel à des experts extérieurs ;

— permettre un recours du législatif à des prestations fournies par les services administratifs à l'instar de ce qui existe en matière d'analyse macroéconomique (prestations de l'INSEE et de la Direction de la prévision pour le Sénat et le Conseil économique et social) ;

— rendre plus systématiques les procédures prévues de saisine de la Cour des comptes par l'Assemblée nationale et le Sénat tout en assurant mieux les fonctions de conseil que la Cour peut remplir auprès du gouvernement.

3) **Le Commissariat général du plan pourrait assurer une mission d'animation et d'information permanente dans le domaine de l'évaluation, en raison de sa vocation interministérielle et de sa position charnière vis-à-vis des acteurs économiques et sociaux.** Il peut tout d'abord remplir des tâches de synthèse et de diffusion méthodologique à l'instar du rôle qu'il tient en matière de calcul économique (production de circulaires de méthodes, comparaisons internationales, etc.). Il peut être un endroit privilégié pour réaliser une confrontation dans le cas de multiévaluations. Par ailleurs, le Commissariat général du plan pourrait se voir, assez naturellement, confier une responsabilité particulière dans l'évaluation des politiques dépassant le cadre habituel, notamment budgétaire, des procédures de décisions ministérielles. Il s'agit d'une part des politiques de moyen terme, maintenues sur longue période, dont les effets sont cumulatifs et globaux (par

exemple sécurité routière, politiques de l'environnement et de l'espace, politiques de restructuration industrielle, politiques de la concurrence). Le deuxième type de politiques dont l'évaluation pourrait être confiée au Commissariat général du plan correspond aux politiques générales pour lesquelles les conditions d'application locales sont importantes. Ce rôle d'évaluation suppose que ne soit pas privilégiée la recherche d'un consensus, traditionnelle dans les travaux du plan, au détriment de l'exercice sans partage d'une responsabilité de jugement sur les résultats.

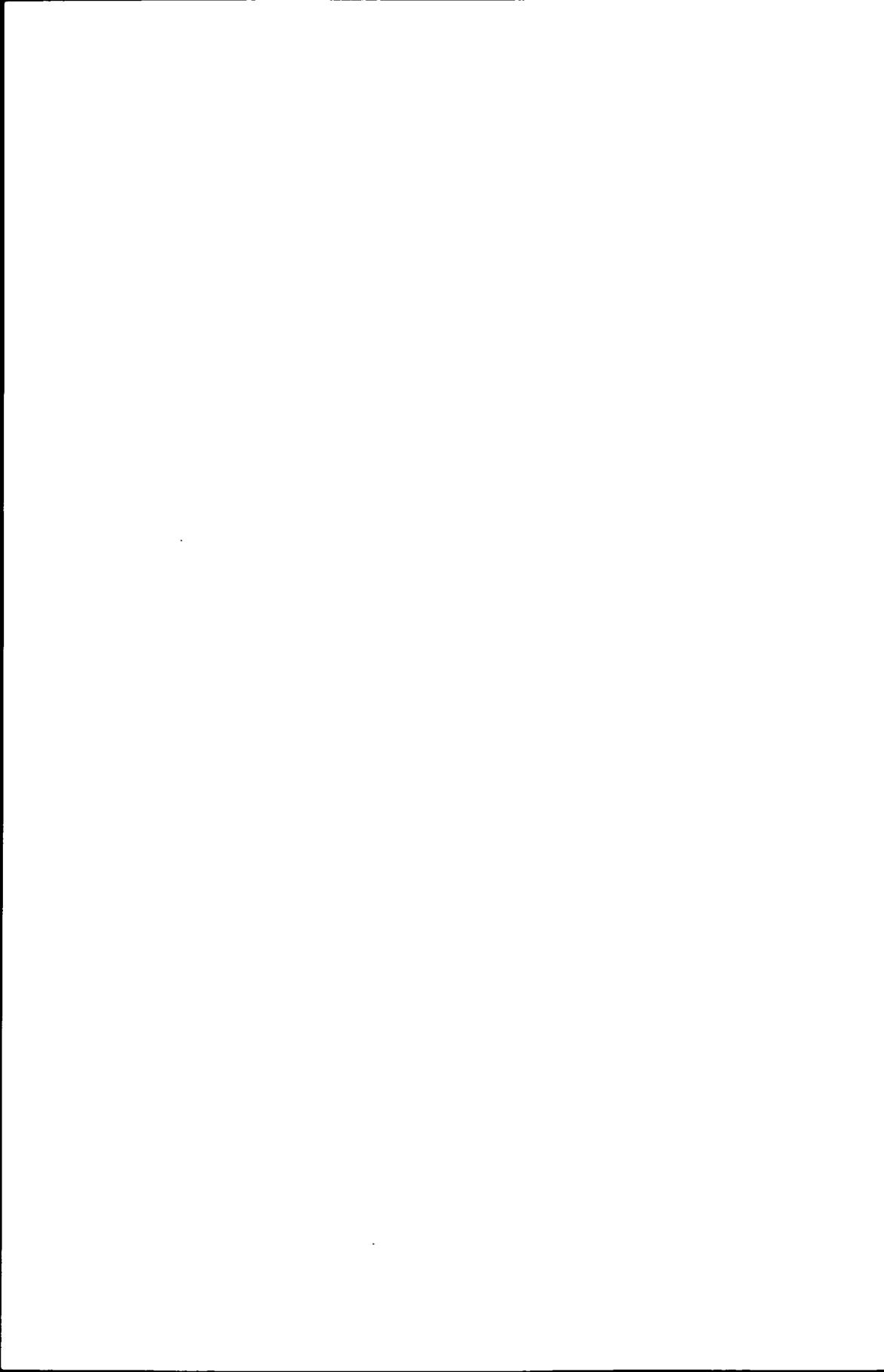
4) Enfin, la diffusion des évaluations par les médias doit être favorisée à partir de pratiques s'inspirant de celles suivies aux États-Unis par le General Accounting Office (conférences de presse, communication de résumés non techniques directement utilisables, etc.).

VI - 5 Identifier les thèmes prioritaires

La relative faiblesse des moyens rapidement utilisables pour réaliser des évaluations solides impose de canaliser l'effort sur quelques thèmes bien choisis importants en termes économiques ou sociaux, et offrant des possibilités à l'évaluateur.

1) L'évaluation doit tout d'abord être utilisée pour **sa fonction de remise en cause**. Une attention particulière doit être portée aux « **politiques silencieuses** » qui n'ont pas d'implication budgétaire directe et tombent plus facilement dans l'oubli sans cesser d'agir. Cette observation concerne prioritairement les **politiques réglementaires**. Dans le même ordre d'idées, il convient de s'intéresser aux « **politiques enracinées** » faisant l'objet, pour ainsi dire, d'une perpétuelle reconduction tacite. Sans aller jusqu'à prôner des pratiques de révision systématique des crédits votés, il convient de s'intéresser aux politiques résultant d'un processus d'empilement long, coûteuses en termes budgétaires. Cette observation concerne tout particulièrement les diverses aides aux entreprises, agricoles comme industrielles, dont l'efficacité sur l'orientation des choix est loin d'être assurée, ainsi que de nombreux transferts sociaux, dont, au contraire, les apparentes vertus redistributrices sont susceptibles de dissimuler des effets incitatifs néfastes.

2) Par ailleurs, au delà de cette problématique d'efficacité et de réduction de coûts implicites ou explicites, l'évaluation doit porter sur les politiques apparaissant particulièrement déterminantes pour le développement sur moyenne période, de l'économie. Tel est le cas en particulier de l'enseignement, de la recherche et de la santé.



Glossaire



Glossaire¹

Analyse des données :

Ensemble des techniques statistiques permettant de traiter de l'information brute (généralement issue d'enquêtes statistiques ou de fichiers de gestion) afin d'en tirer les enseignements utiles.

Analyse multivariée :

Ensemble de techniques statistiques reposant sur des relations linéaires et permettant de juger de la présence ou de l'absence d'un effet.

Dans le cas de l'analyse de la variance, les variables explicatives sont des variables artificielles prenant les valeurs 0 ou 1, qui indiquent la présence ou l'absence d'un effet sur une variable.

Dans le cas de l'analyse de la covariance, les variables explicatives sont d'une part des variables numériques, d'autre part des variables 0 et 1.

Données en coupes transversales :

Observations relatives à une même unité de temps (mois, trimestre, année) et collectées sur des sujets ou des événements différents.

Données en séries temporelles :

Observations collectées sur un nombre plus ou moins grand d'unités de temps successives, mais concernant constamment le même sujet (ou le même ensemble de sujets).

Données sur panel :

Mélange d'observations en coupes transversales et en séries temporelles (on repère l'évolution au cours du temps des caractéristiques de chacun des sujets observés).

Econométrie :

Ensemble des techniques statistiques permettant d'estimer numériquement les paramètres d'une relation analytique entre une variable expliquée et des variables explicatives, cette relation étant postulée en général à partir d'une théorie *a priori* (notamment économique).

Etude de cas :

Description analytique d'un événement, d'un organisme, d'un processus ou d'un programme.

Evaluateur :

Personne(s), groupe de personnes ou d'organismes ayant la charge de réaliser une évaluation.

1. Ce glossaire constitue une aide à la lecture du rapport en recensant les principaux termes techniques utilisés. Les définitions données se veulent pédagogiques plus que totalement rigoureuses.

Evaluation externe :

Cas où l'évaluateur est indépendant du (ou des) organisme(s) qui a (ont) décidé ou mis en oeuvre la politique.

Evaluateur interne :

Cas où l'évaluateur dépend du (ou des) organisme(s) qui a (ont) décidé ou mis en oeuvre la politique.

Expérimentation vraie :

Attribution de traitements (ou de politiques) différents(es) à des groupes représentatifs de la politique étudiée et équivalents (ce qui implique qu'ils résultent d'un tirage au sort).

Groupes équivalents :

Groupes (d'individus, d'entreprises, etc.) ayant la même distribution de caractéristiques.

Groupes non équivalents :

Groupes (d'individus, d'entreprises, etc.) dont les distributions de caractéristiques sont différentes.

Groupes non équivalents issus de l'observation :

Groupes dont la constitution et le traitement ne résultent pas de la volonté de l'évaluateur, mais des conditions effectives d'application de la politique à évaluer.

Groupes non équivalents issus d'une quasi-expérimentation :

Groupes dont le traitement résulte de la volonté de l'évaluateur mais qui, n'étant pas tirés au sort, ne sont pas strictement équivalents.

Groupe témoin :

Groupe (d'individus, d'entreprises, etc.) n'ayant pas subi la politique que l'on cherche à évaluer.

Groupe de traitement :

Groupe (d'individus, d'entreprises, etc.) auquel l'on applique la politique que l'on veut évaluer.

Mesure :

Procédure consistant à assigner un nombre à une caractéristique donnée d'un objet ou d'un événement.

Modélisation :

Construction d'une représentation schématique formalisée de comportements et de mécanismes à partir d'une théorie *a priori* explicative, et susceptible d'estimation numérique.

Plan de recherche (ou protocole) :

Organisation méthodologique générale de l'évaluation en fonction de l'information empirique (observations) disponible ou à collecter.

Prescripteur (d'une évaluation) :

Personne(s), groupe de personnes ou d'organismes demandant que la politique publique soit évaluée, et en confiant la mission à l'évaluateur.

Quasi-expérimentation :

Constitution, par l'évaluateur, de groupes de comparaison par un procédé autre que le tirage au sort, mais qui vise à limiter autant que possible les différences de caractéristiques entre ces groupes.

Typologie :

Regroupement d'observations en classes possédant des caractéristiques aussi homogènes que possible.

Technique d'analyse :

Outil particulier de traitement d'une information ou d'un ensemble d'informations.

Validité conceptuelle :

Légitimité de généraliser les conclusions de l'évaluation à d'autres variantes de la même politique.

Validité externe :

Légitimité de généraliser les conclusions de l'évaluation à la même politique visant d'autres cibles ou opérant dans un autre contexte.

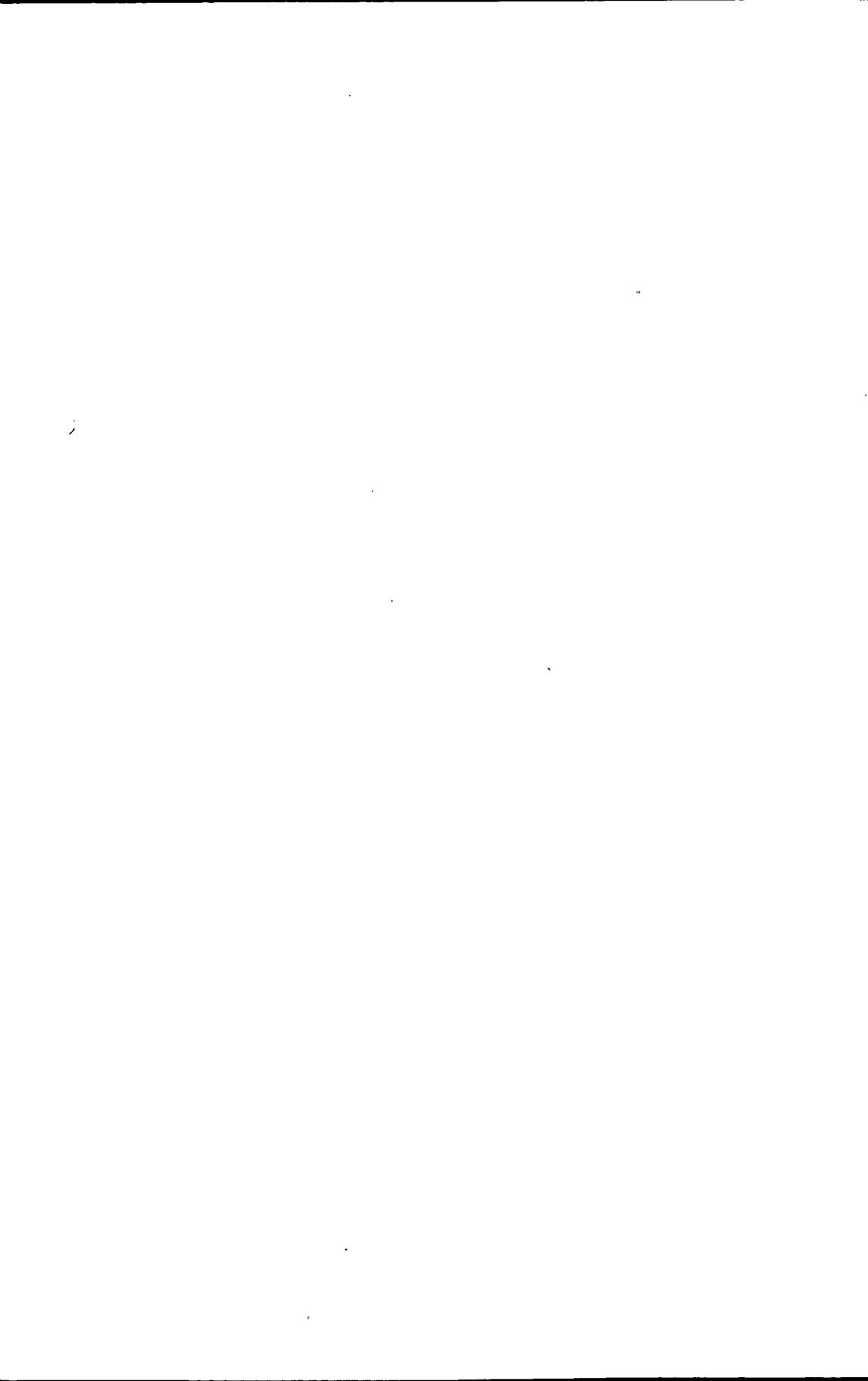
Validité interne :

D'une part, possibilité de déceler des corrélations entre la politique évaluée et certains phénomènes. D'autre part, légitimité de considérer ces phénomènes comme des effets de la politique.

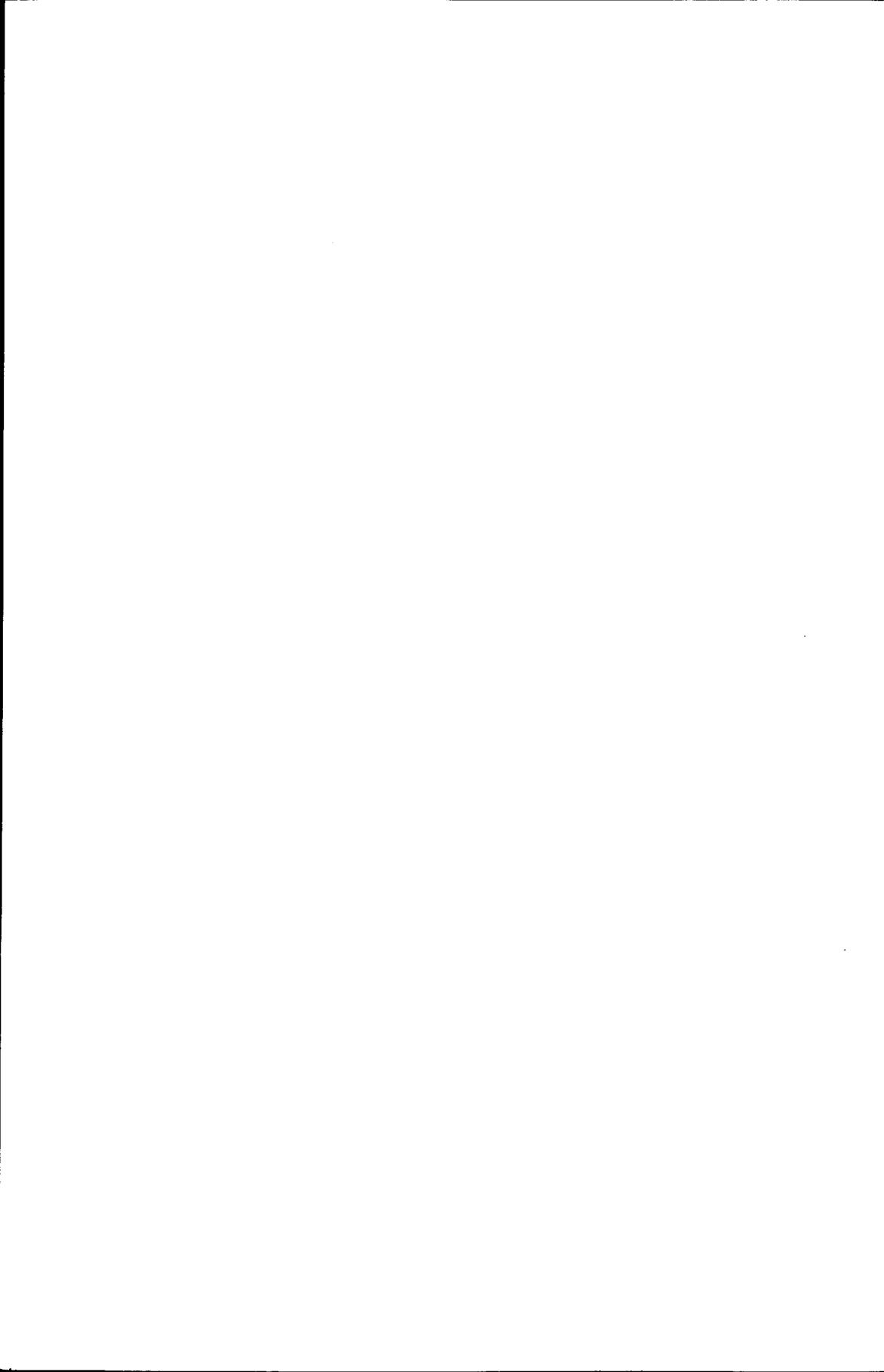
Variance :

Mesure de la dispersion des caractéristiques d'une population.

Créer de la variance : multiplier les observations d'unités suffisamment dissemblables au sein d'une population et mesurer leurs caractéristiques.



Annexes



Annexes

- 1 — Composition du groupe de travail au 30/12/85
- 2 — Communications présentées au groupe

Composition du groupe de travail « Méthodes d'évaluation des politiques publiques » au 30/12/85

Président

M. Michel Deleau Sous-directeur à la Direction de la prévision. Ministère de l'Économie, des Finances et du Budget

Secrétariat du groupe :

M. Philippe Penz Chargé de mission au Service économique du Commissariat général du plan

Membres de la cellule d'animation

M. Jean-Pierre Nioche Professeur à l'École des hautes études commerciales
M. Robert Poinard Secrétaire de la Commission RCB - Direction de la prévision - Ministère de l'Économie, des Finances et du Budget

Membres permanents :

M. Thierry Bert (*) Inspecteur des finances, chargé de mission auprès du chef de l'Inspection des finances

M. Bernard Bobe Conseiller scientifique au Commissariat général du plan

Mme Aude-Marie Debarle Chargée de mission au Service des études et de la recherche du Commissariat général du plan

M. Bernard Dechambre Chargé de mission à la Mission d'analyse et d'évaluation - Direction des Affaires économiques et financières - Ministère de l'Agriculture

M. Bernard Devev Directeur à la COFREMCA

Mme Elvyne Février Chargée de mission auprès du Sous-directeur de la sécurité et de l'exploitation de la route - Direction de la sécurité et de la circulation routière - Ministère de l'Urbanisme, du Logement et des Transports

(*) M. Thierry Bert a succédé à M. Laurent Perpère, inspecteur des Finances.

M. Thierry Gaudin	Directeur du Centre de prospective et d'évaluation - Ministère de la Recherche et de l'Industrie
M. Daniel Goudard	Chargé de mission au Commissariat général du plan - Adjoint au chef du Service économique
M. Pierre Greiner (*)	Chef de la Mission d'analyse et d'évaluation - Direction des affaires économiques et financières - Ministère de l'Agriculture
M. Christian Join- Lambert(**)	Conseiller référendaire à la Cour des comptes
M. Pierre Mondon	Chef du Service de la prévision, des statistiques et de l'évaluation (SPRESE) - Ministère de l'Éducation Nationale
M. Henry Ollagnon	Chargé de mission à la Mission d'analyse et d'évaluation - Direction des affaires économiques et financières - Ministère de l'Agriculture
Mme Elisabeth Pascaud (***)	Responsable d'Études au Centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC)
Mme Brigitte Roguet	Responsable du secrétariat des comptes de l'emploi - Service des études et de la statistique - Ministère du Travail, de l'Emploi et de la Formation professionnelle
M. Daniel Schwartz	Directeur de l'unité de recherches statistiques de l'Institut national de la santé et de la recherche médicale (INSERM)
M. Olivier Villey	Chef de la division de l'emploi et des politiques de l'emploi - Service des études et de la statistique - Ministère du Travail, de l'Emploi et de la Formation professionnelle
M. Michel Volle	Chef de la Mission d'études économiques - Centre national d'études des télécommunications

(*) M. Pierre Greiner a succédé à M. Dominique Gagey.

(**) M. Christian Join-Lambert a succédé à M. Jean-Marie Bertrand, conseiller référendaire à la Cour des comptes.

(***) Mme Elisabeth Pascaud a succédé à Mme Catherine Blum, directeur du CREDOC.

Communications présentées au groupe :

Les méthodes de l'évaluation (31/01/1985)

- Repérage de l'évaluation par rapport à des genres voisins. Aperçu des problèmes méthodologiques de l'évaluation (M. Nioche, HEC).
- Les techniques d'analyse causale et de mesure des effets globaux (M. Volle, Centre national d'études sur les télécommunications).
- Évaluation et expérimentation : le cas de l'épidémiologie (M. Schwartz, INSERM).
- Évaluation et calcul économique (M. Henry, Laboratoire d'économétrie de l'École polytechnique).

Les méthodes d'évaluation en matière de politique d'éducation (14/02/1985)

- L'observatoire des entrées dans la vie active (M. Germe, Centre d'études et de recherches sur l'emploi et les qualifications).
- L'Évaluation en matière pédagogique au ministère de l'Éducation nationale (M. Tournardre, Service de la prévision, des statistiques et de l'évaluation, ministère de l'Éducation nationale).
- Une évaluation internationale de l'enseignement des mathématiques (M. Robin, Institut national de recherche pédagogique, ministère de l'Éducation nationale).

Les méthodes d'évaluation en matière de politiques sociales (07/03/1985)

- La problématique générale de l'évaluation en matière de politiques sociales (Mme Catherine Blum, CREDOC).
- Mères isolées, allocation de parent isolé, et travail. Essais de mesure de l'efficacité économique et sociale de l'API (M. Ray, université de Nancy II, LASARE).

Les méthodes d'évaluation en matière de politiques de transports et d'environnement (21/03/1985)

- La problématique générale de l'évaluation en matière de transports (M. Gabet, Direction des transports terrestres, ministère de l'Urbanisme, du Logement et des Transports).
- L'évaluation dans le domaine de la sécurité routière (Mme Février, Direction de la sécurité et de la circulation routières, ministère de l'Urbanisme, du Logement et des Transports).
- L'inventaire permanent du littoral, outil d'évaluation en matière de politique d'aménagement (M. Ecobichon, Direction de l'urbanisme et des paysages, ministère de l'Urbanisme, du Logement et des Transports).

L'évaluation du plan textile (30/04/1985)

- Exposé de M. Tuloup (Direction des industries chimiques, textiles et diverses, ministère du Redéploiement industriel et du Commerce extérieur).

- Exposé de MM. Boaretto et du Buisson (Direction de la prévision et Direction du budget, ministère de l'Économie, des Finances et du Budget).
- Les méthodes utilisées pour l'évaluation du plan textile (M. Nioche, HEC).

Les méthodes de l'évaluation en matière de politique de santé (04/06/1985)

- La problématique générale de l'évaluation en matière de politique de santé (Mme Weill, Direction des hôpitaux, ministère des Affaires sociales et de la Santé).
- L'évaluation des innovations technologiques (Mme Claudine Blum, Direction du plan de l'Assistance publique de Paris).

Les méthodes d'évaluation des politiques de la recherche (18/09/1985)

- L'évaluation des Directions régionales de l'industrie et de la recherche (M. Gaudin, Centre de prospective et d'évaluation, ministère de la Recherche et de l'Industrie).
- Le Comité national d'évaluation des universités (M. Staropoli, Comité national d'évaluation des universités, ministère de l'Éducation nationale).

L'évaluation par la société civile (03/10/1985)

- Le rôle des médias dans le déclenchement et dans la publicité de l'évaluation (M. de La Taille, TF1).
- Évaluation par les citoyens : apport des systèmes d'observation et de mesure des sensibilités et des mentalités du public (M. Deveze, Cofremca).

Les méthodes d'évaluation des politiques d'emploi (17/10/1985)

- Les méthodes d'évaluation en matière de politiques d'emploi (Mme Elbaum, délégation à l'Emploi).
- La mise en place de l'évaluation des travaux d'utilité collective (M. Villey, Direction des études statistiques, ministère du Travail, de l'Emploi et de la Formation professionnelle).
- Les méthodes d'évaluation utilisées dans l'étude prioritaire RCB sur les aides à l'emploi (M. Hara, Direction de la prévision, ministère de l'Économie, des Finances et du Budget).
- L'apport des enquêtes auprès des entreprises dans l'évaluation des aides à l'emploi (M. Eymard-Duvernet, directeur du Centre d'études de l'emploi).
- Les modèles micro-analytiques de fonctionnement du marché du travail et leur utilité pour l'évaluation des politiques d'emploi (M. Ballot, université Paris II).

Les méthodes d'évaluation des politiques agricoles (24/10/1985)

- La pratique de l'évaluation au ministère de l'Agriculture (M. Gagey, Mission d'analyse et d'évaluation, ministère de l'Agriculture).
- Présentation méthodologique de l'étude d'évaluation de la politique publique de l'élevage bovin (M. Le Pen, Legos, université Paris IX).

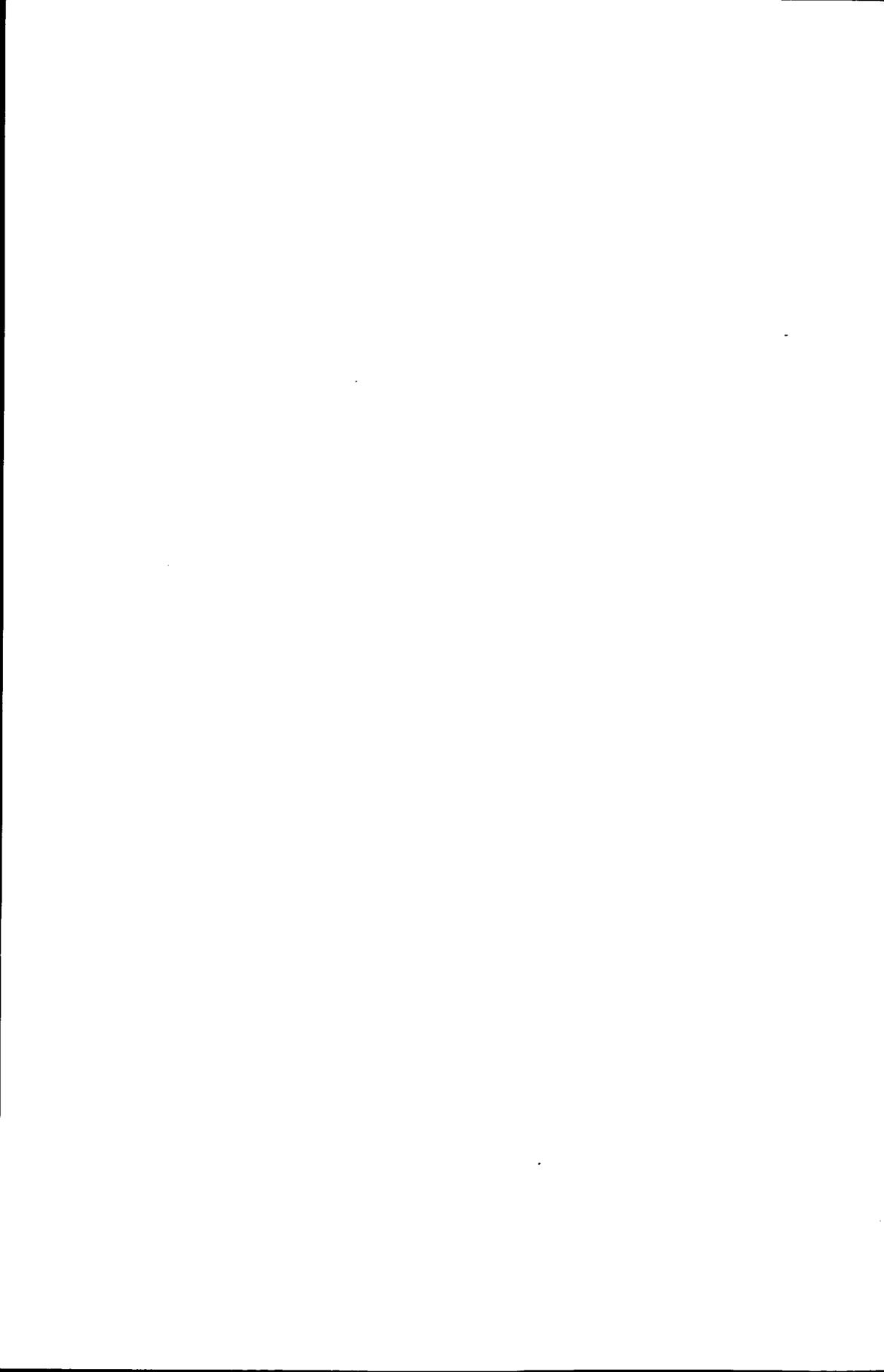
- L'approche « Système-acteurs » (M. Ollagnon, Mission d'analyse et d'évaluation, ministère de l'Agriculture).

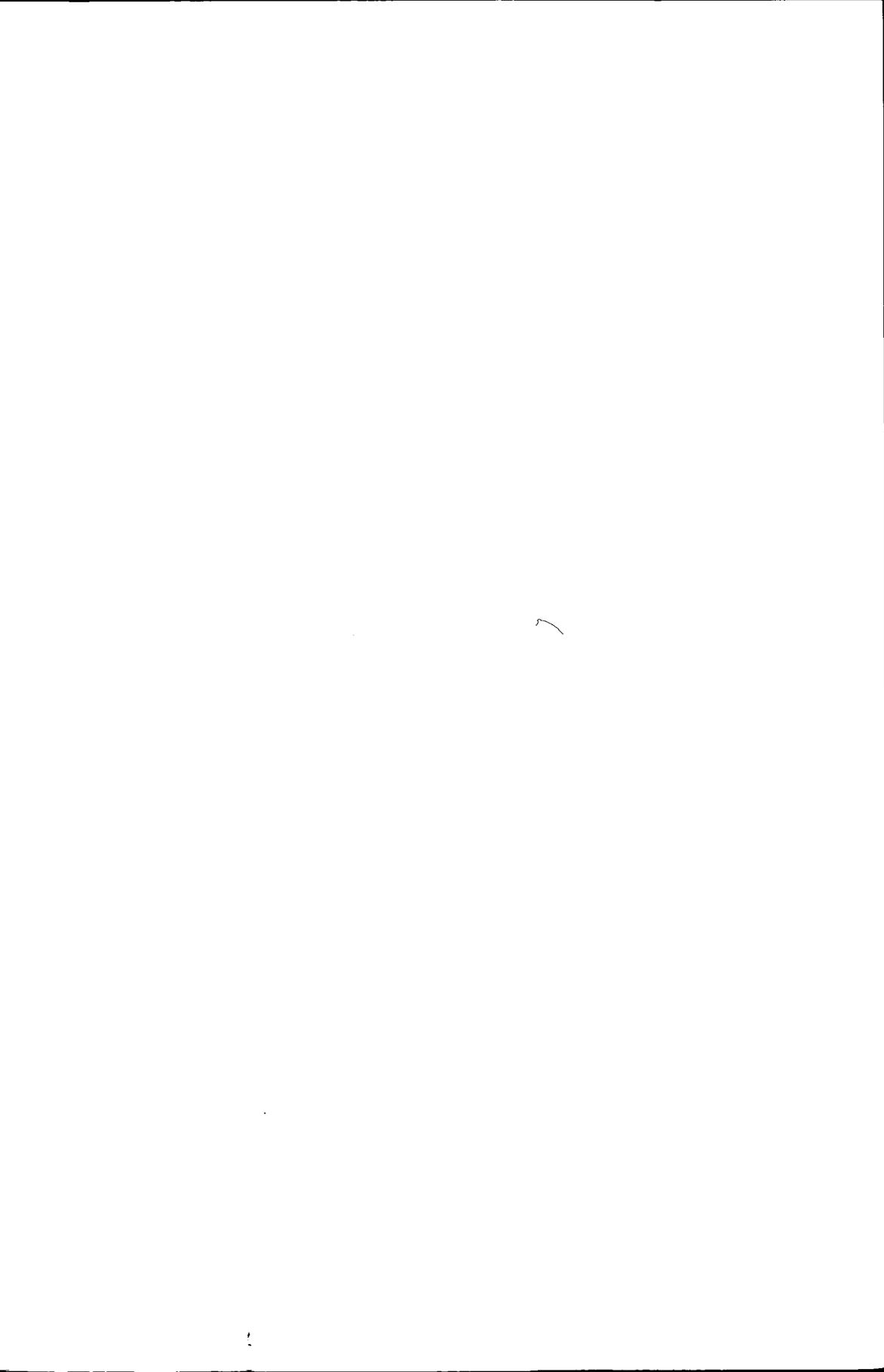
Contrôle et évaluation (07/11/1985)

- Les problèmes d'organisation générale des rapports entre contrôle et évaluation (M. Joint-Lambert, Cour des comptes).
- Le rôle de l'Inspection des finances (M. Bert, Inspection des finances).
- Le rôle des inspections générales ministérielles (M. Milloz, Inspection générale de l'industrie).
- Le rôle de la Cour des comptes (M. Logerot, secrétaire général de la Cour des comptes).
- Les corps de contrôle et l'évaluation (M. Gibert, CESMAP).

Les expériences étrangères en matière d'évaluation (11/12/1985)

- L'évaluation des politiques publiques aux Pays-Bas (M. Sorber, ministère des Finances des Pays-Bas).
- L'évaluation des politiques publiques aux États-Unis (M. Rouban, Centre de recherches administratives et politiques de l'université de Rennes I).



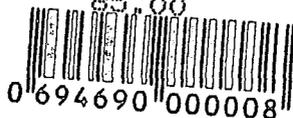




EVALUER POLIT PUBLI

85,00

504



LA DOCUMENTATION FRANÇAISE
29-31, quai Voltaire
75340 Paris Cedex 07
Tél. : (1) 42.61.50.10
Télex : 204826 Docfran Paris

Imprimé en France
ISBN : 2-11-001702-3
DF 1345