

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 Poinard

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.

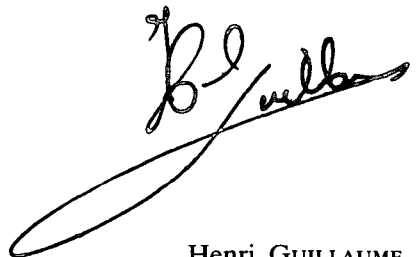
De manière plus précise, l'examen des expériences étrangères et françaises les plus significatives doit avoir pour fonction de recenser les causes de succès et d'échec des évaluations menées et d'apprécier les résultats atteints. Il visera à retirer des enseignements généraux sur les conditions de faisabilité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

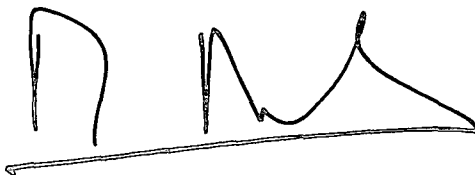
Le rapport n'avait pas comme objectif de fixer dans le détail, notamment institutionnel, la mise en œuvre de ces diverses recommandations. Elles ont paru, aux membres du groupe, susceptibles de garantir un développement solide et pluraliste de l'évaluation des politiques publiques, adapté à l'environnement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

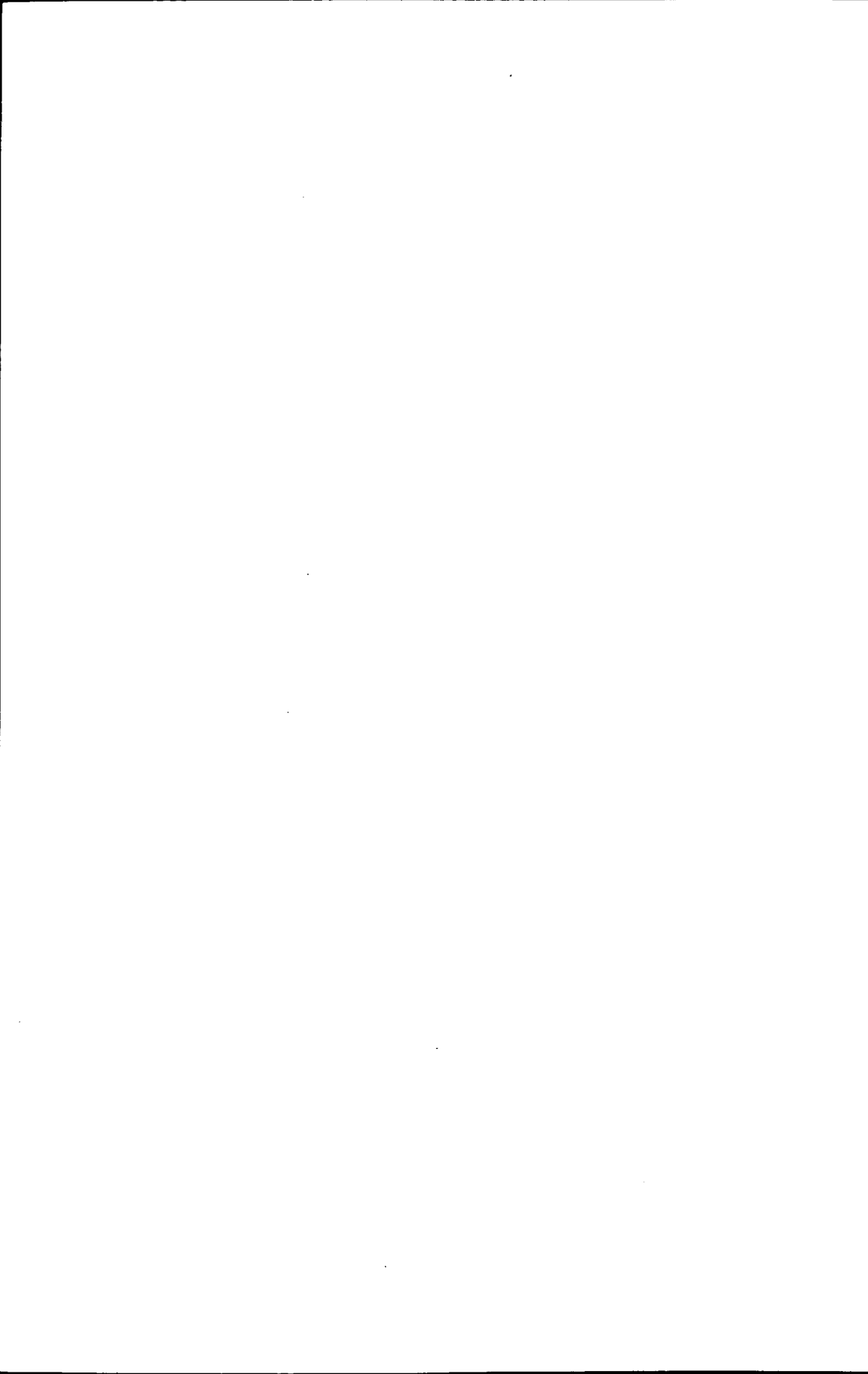
La diversité et la qualité de ces diverses contributions, dont je remercie ici les auteurs, me paraît également porter témoignage des chances existant actuellement pour un développement de l'évaluation des politiques publiques en France. Je souhaite que ce rapport, réalisé sous l'égide du Commissariat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

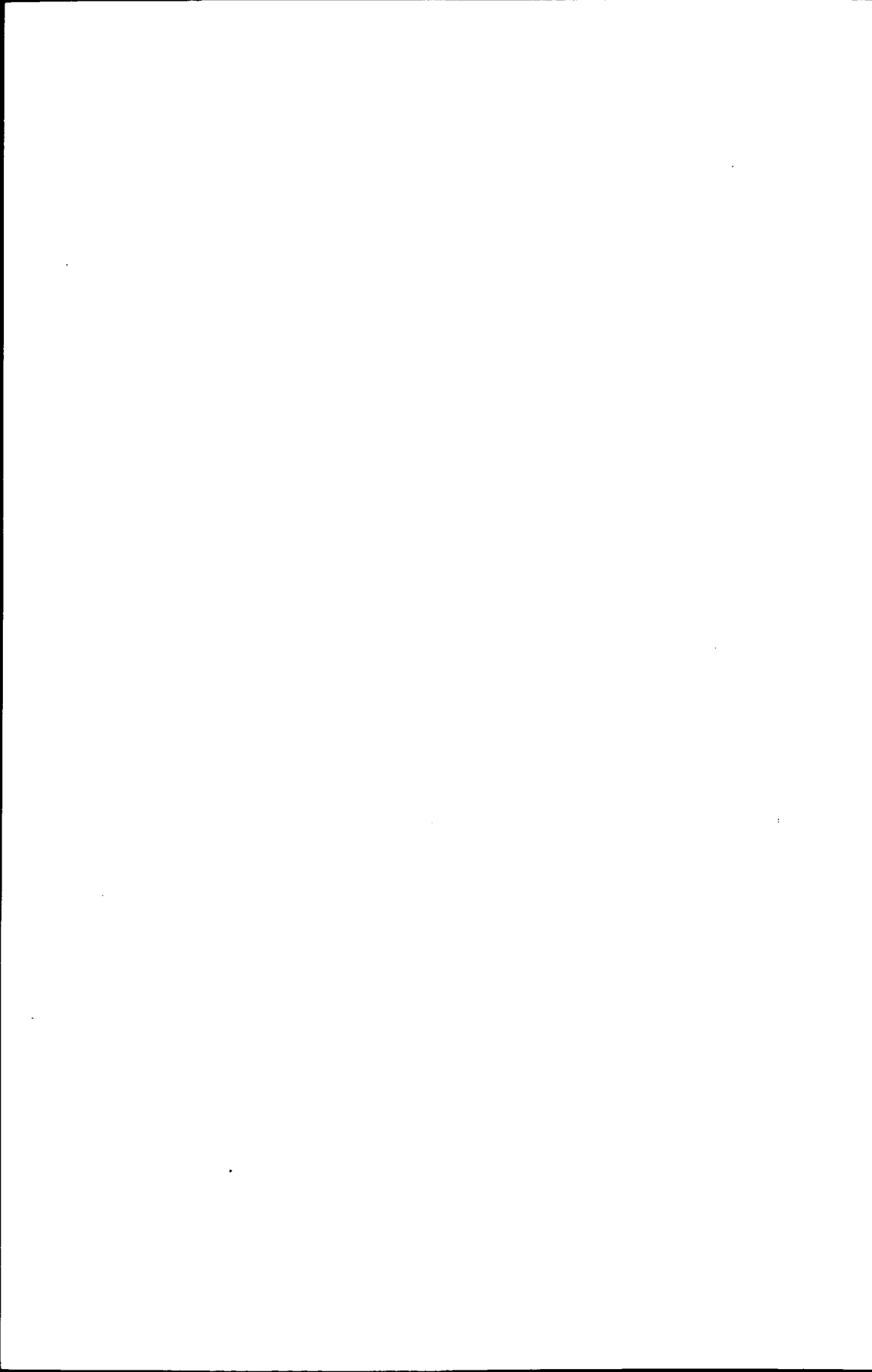
VI — Développer l'évaluation des politiques publiques en France : cinq propositions

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).

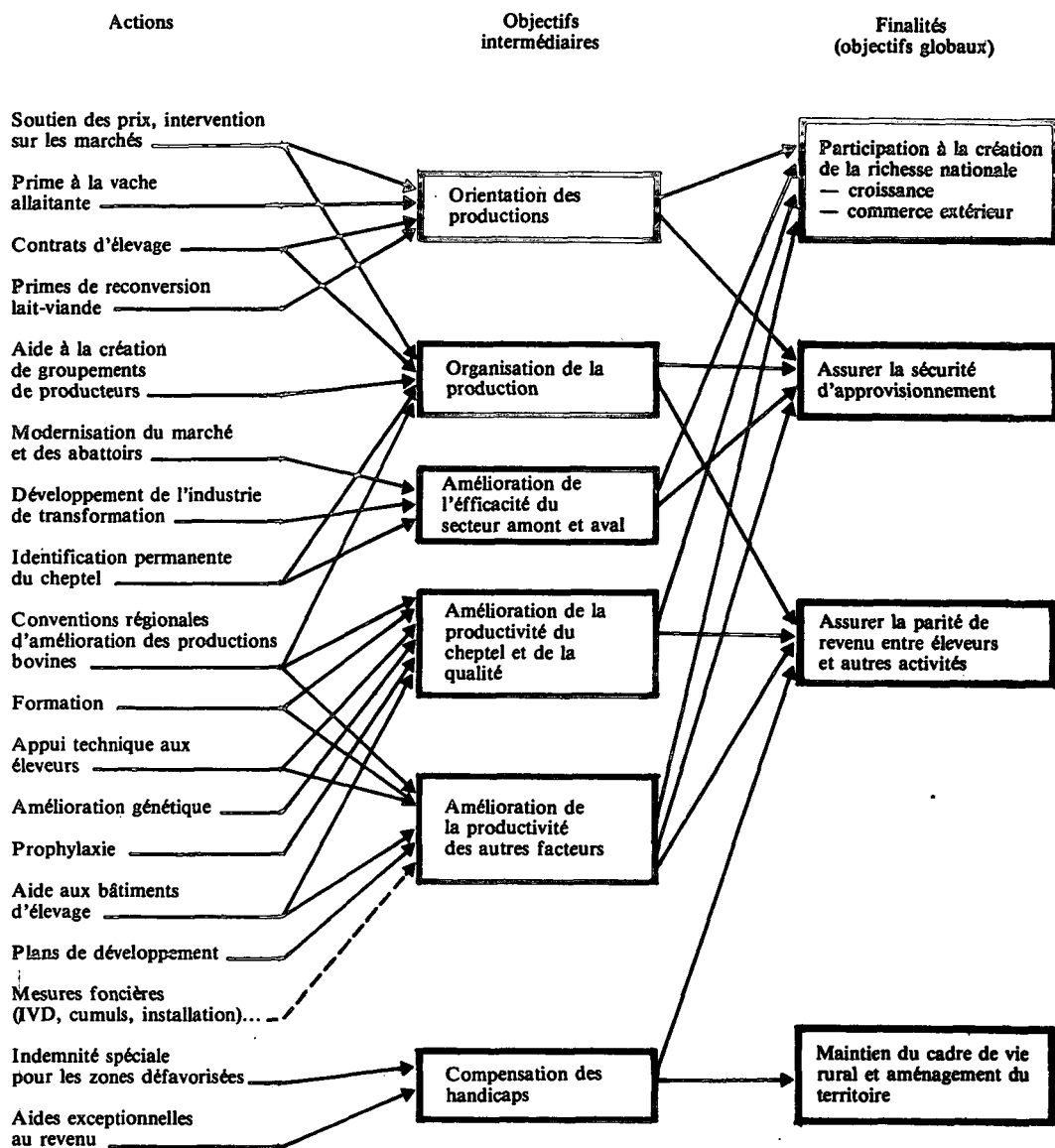
Une fois reconnu « l'objet » de l'évaluation, il convient d'identifier les **objectifs** poursuivis par la politique. Nous reviendrons sur ce thème ultérieurement, en distinguant entre **objectifs finaux**, souvent qualitatifs et généraux, et **objectifs intermédiaires** (ou opérationnels) traduisibles 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.

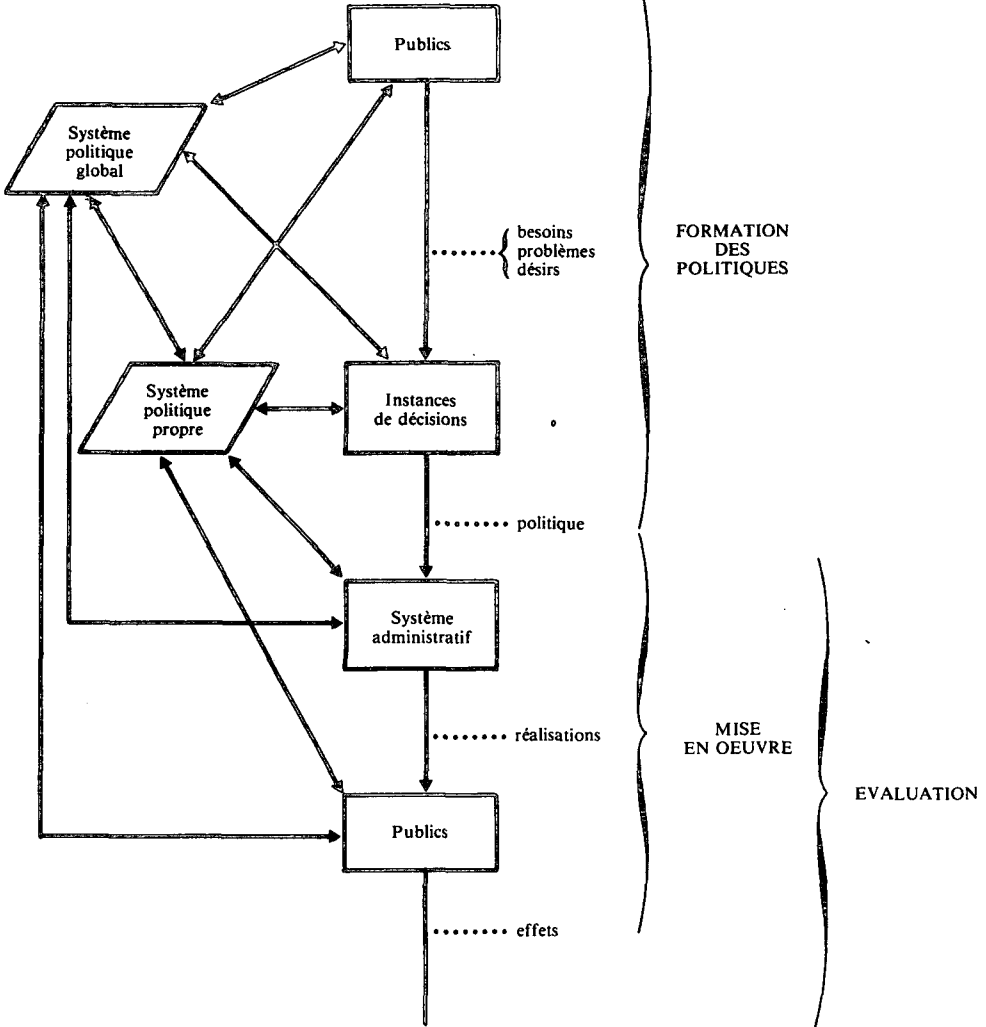
Enfin, l'analyse des politiques publiques a généralement une ambition explicative et pas seulement descriptive. Elle suppose une conceptualisation des processus en cause et la définition d'une méthodologie permettant de générer et de tester des hypothèses causales. Les méthodologies sont variables et 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 efficacité ? ») : 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

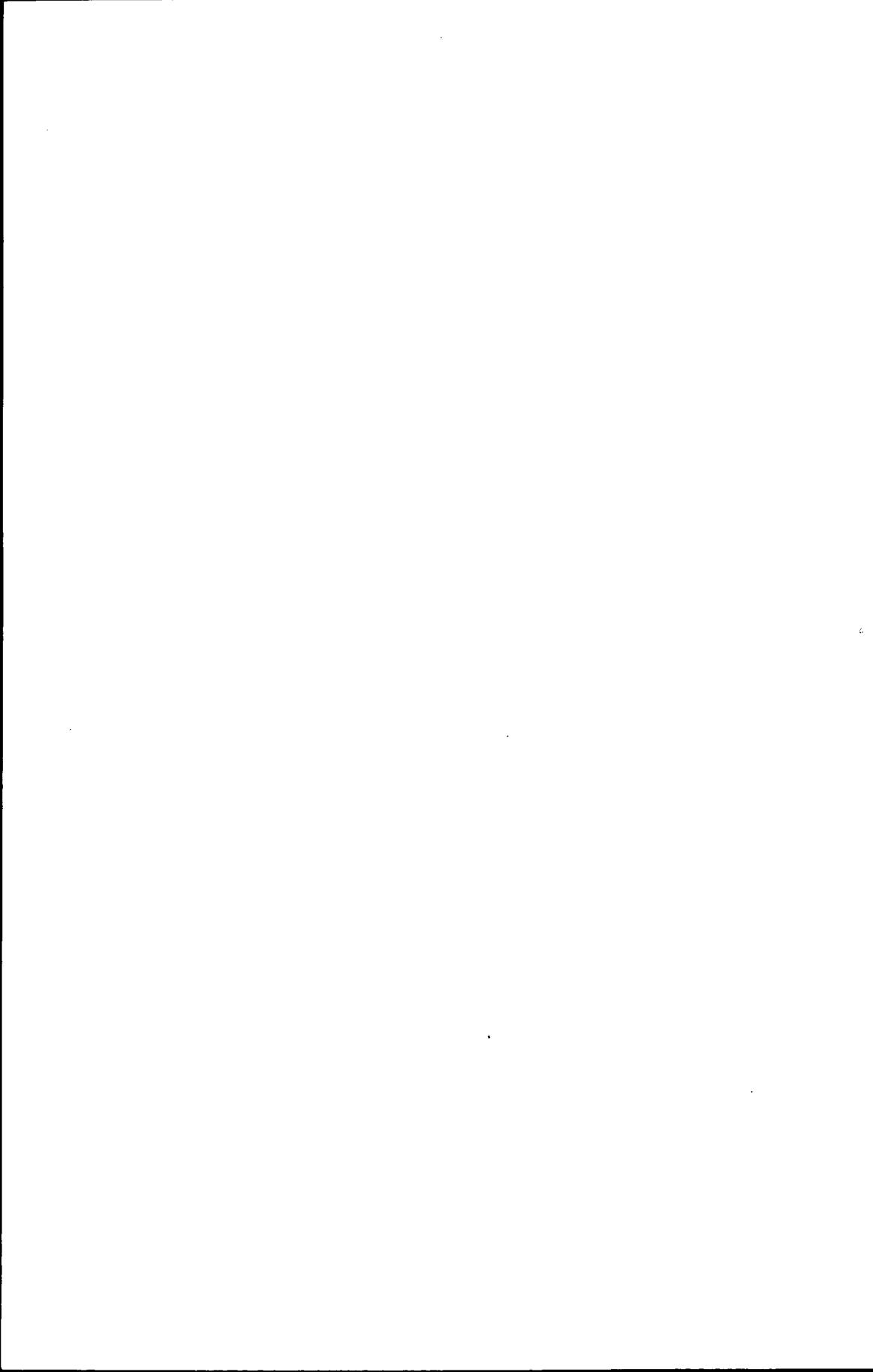
pertinente ou, à tout le moins, constituer un appui très utile³². De même, en ce qui concerne le cycle politique, qu'on passe de l'*ex post* proprement dit à des problèmes de formulation de politique, alors les démarches associatives retrouvent leurs droits.

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.

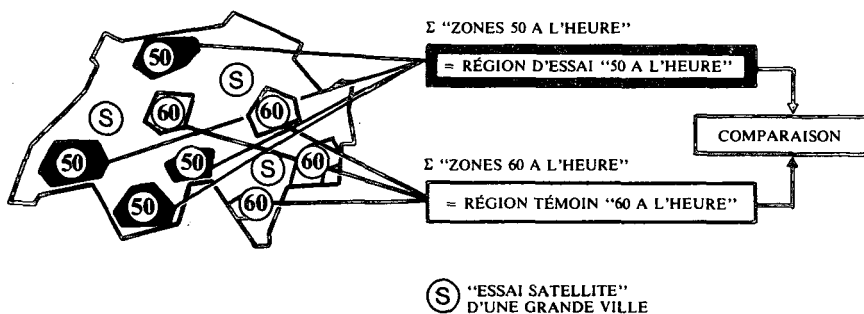
24. L'analyse multivariée a notamment été utilisée pour l'insertion des 16-18 ans (cf. encadré n° 6). Les conclusions auxquelles elle a conduit dans ce cas sont toutefois controversées, d'une part parce que les groupes ne sont pas réellement équivalents et d'autre part en raison du choix des variables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

