

méthodes
d'*évaluation*
des politiques publiques
actes du séminaire

MÉTHODES D'ÉVALUATION DES POLITIQUES PUBLIQUES

actes du séminaire

Sous la direction de

Nicolas Studer

Directrice de la publication

Anne-Marie Brocas

Interventions de

- Annie Fouquet** (Société française d'évaluation), p. 9 ;
Esther Duflo (Collège de France), p. 17 ;
Bruno Crépon (CREST-INSEE), p. 39 ;
Pauline Givord (INSEE), p. 49 ;
Jean-Claude Moisdon (ENSMP), p. 61 ;
Hélène Genin (Euréval), p. 71 ;
Francine Fenet et Audrey Daniel (ville de Nantes), p. 85 ;
Sylvain Ortis (région Rhône-Alpes), p. 95 ;
Jean-René Brunetière (CGEDD), p. 113 ;
Isabelle Engsted-Maquet (Commission européenne), p. 123.

Illustrations fournies par les intervenants

Coordination éditoriale

Élisabeth Hini

« En application de la loi du 11 mars 1957 (art. 4) et du code de la propriété intellectuelle du 1^{er} juillet 1992, toute reproduction totale ou partielle à usage collectif de la présente publication est strictement interdite sans autorisation expresse de l'éditeur. Il est rappelé à cet égard que l'usage abusif et collectif de la photocopie met en danger l'équilibre économique des circuits du livre. »

Sommaire

MÉTHODES D'ÉVALUATION DES POLIQUES PUBLIQUES

Introduction.....	5
Définitions de l'évaluation	7
L'évaluation des politiques publiques en France. Définitions et historique	9
L'évaluation d'impact en économie du développement. L'approche expérimentale.....	17
Synthèse des débats avec la salle.....	29
Bibliographie	34
L'identification de l'effet propre d'un programme	37
Le cadre général d'analyse, les méthodes d'appariement et les expériences contrôlées	39
Les méthodes de différences de différences, de variables instrumentales et de régressions sur discontinuités	49
Bibliographie	57
Évaluation et changements organisationnels :	
l'apport des études de terrain	59
L'approche par la recherche-intervention	61
L'apport de l'approche monographique	71
Synthèse des débats avec la salle	78
Bibliographie	82
Associer les usagers et les citoyens	83
L'importance de la prise en compte des besoins et des attentes de la population dans les évaluations	85
L'association des bénéficiaires et des citoyens au processus d'évaluation des politiques régionales	95
Synthèse des débats avec la salle	105
Bibliographie	109
L'approche objectifs/indicateurs	111
Les indicateurs de la LOLF. Une occasion de débat démocratique?	113
L'approche objectifs/indicateurs pour l'évaluation des politiques sociales : le cadre européen	123
Synthèse des débats avec la salle	134
Bibliographie	139

Introduction

Le décret du 22 janvier 1990 a instauré la procédure interministérielle d'évaluation. L'évaluation consiste à « rechercher si les moyens juridiques, administratifs ou financiers mis en œuvre permettent de produire les effets attendus d'une politique et d'atteindre les objectifs qui lui sont fixés ». Au-delà de cette définition institutionnelle, l'évaluation s'est imposée comme un domaine à part entière des sciences sociales, recouvrant notamment la recherche évaluative, qui vise à identifier les effets propres d'une politique, en les distinguant des effets de contexte ou produits par d'autres politiques menées concomitamment. L'évaluation contribue aussi à améliorer la politique publique elle-même.

En tant que service statistique ministériel, la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) a pour vocation d'occuper une place centrale dans les dispositifs d'évaluation des politiques sanitaires et sociales. Elle produit tout d'abord un certain nombre de données qui peuvent être mobilisées par différents acteurs institutionnels ou par des chercheurs pour évaluer les politiques. Elle peut également être amenée à effectuer elle-même un certain nombre d'études évaluatives permettant d'éclairer le débat public. Enfin, elle participe aux différentes instances d'évaluation mises en place dans son champ de compétences.

En vue de promouvoir la démarche d'évaluation dans le champ des politiques sanitaires et sociales et favoriser l'usage des outils appropriés pour conduire ces évaluations, la DREES a organisé, au cours de l'année 2009-2010, un séminaire sur les méthodes de l'évaluation *ex post*, à savoir celle qui intervient après la mise en œuvre d'un dispositif particulier. L'évaluation *ex-ante*, qui mesure l'impact attendu d'une politique, était exclue de ce champ, de même que les évaluations qui interviennent au fur et à mesure que se met en place une politique. Ce séminaire n'a pas eu non plus vocation à couvrir les aspects institutionnels visant à gérer l'évaluation. Il était centré sur la présentation des outils de l'économie et de la sociologie pouvant contribuer à l'évaluation.

Chaque séance de ce séminaire construite autour de deux présentations exposait une approche particulière de l'évaluation, suivie d'un échange avec le public. La première séance s'est attachée à définir précisément l'évaluation. Les trois séances suivantes ont abordé successivement : les méthodes économétriques visant à identifier les effets propres d'une politique ; les changements organisationnels, tels qu'ils sont perçus à travers les études de terrain ; et les méthodes qui permettent d'impliquer les usagers, par l'intermédiaire de *focus groups* et d'enquêtes de satisfaction. La dernière séance s'est enfin intéressée aux approches consistant à associer objectifs, indicateurs et *benchmarking*.

L'objet de ce recueil est de retranscrire le plus fidèlement possible les différents exposés de ce séminaire ainsi que les échanges qui les ont suivis. Nous tenons à remercier une nouvelle fois les intervenants sans qui ce séminaire n'aurait pas pu avoir lieu, ainsi que l'ensemble des participants pour la qualité des échanges et contributions apportées au débat.

Nicolas Studer

DÉFINITIONS DE L'ÉVALUATION

La première séance du séminaire sur les méthodes d'évaluation de politiques publiques était organisée autour de deux interventions :

- une intervention d'Annie Fouquet, présidente de la Société française de l'évaluation, qui a rappelé les concepts de base de l'évaluation et le contexte historique de son développement, notamment institutionnel en France ;
- une intervention d'Esther Duflo, professeur au MIT, titulaire de la chaire « Savoirs contre pauvreté » du Collège de France, qui a présenté l'évaluation d'impact en économie du développement, ce domaine de l'économie apparaissant souvent comme un domaine précurseur en matière d'évaluation. Elle a plus précisément présenté les concepts de base de la recherche évaluative (contrefactuel, groupe de contrôle et de traitement, effet propre...) ainsi que des exemples d'évaluation en économie du développement. Les développements de l'évaluation et les débats autour des évaluations dans le domaine de l'économie du développement ont été également abordés.

Ces interventions ont été suivies d'un débat avec la salle.

L'évaluation des politiques publiques en France

Définitions et historique

Annie Fouquet (Société française de l'évaluation)

Les concepts de base de l'évaluation

L'évaluation de l'action publique (politique, programme, dispositif) a pour objet d'en apprécier la valeur au regard de certains critères, comme la pertinence, l'efficacité, l'efficience..., notamment dans le but d'apporter des connaissances pour aider à la décision et éclairer le débat public.

Une démarche collective...

Le terme d'évaluation est aujourd'hui en vogue et est sujet à bien des malentendus. Chez certains, il fait ressurgir des peurs anciennes, remontant aux temps de l'école et des « zéros pointés ». L'évaluation est perçue par ceux-ci comme un jugement, parfois injuste et qui met en cause des acteurs personnellement. Pour d'autres, notamment les ingénieurs, les physiciens, l'évaluation est synonyme d'estimation. Elle consiste à estimer la valeur d'un paramètre d'un modèle pour estimer l'effet d'une action. Pour autant, estimer la valeur d'un paramètre ne suffit pas à apprécier la valeur de l'action publique, qui ne se résume pas au modèle testé. Économistes et statisticiens ont ainsi tendance à considérer ces travaux comme des évaluations de politiques, alors qu'ils ne sont qu'un élément de constat parmi d'autres à prendre en compte dans le cadre d'une évaluation, comme on le verra plus loin.

On parle également d'évaluation pour les entretiens individuels annuels : en effet dans leur principe, ils s'appuient sur une évaluation de l'action collective dans laquelle s'inscrit l'action individuelle et qui lui donne sens.

L'évaluation dont il est question ici vise à apprécier la valeur d'une action publique. Or, selon sa place (décideur politique, opérateur, bénéficiaire), chacun peut avoir un point de vue différent sur l'action menée. Ainsi il est possible de développer plusieurs jugements de valeur sur la même action. Évaluer une politique menée au nom de l'intérêt général, c'est prendre en compte l'ensemble de ces points de vue. Dans son article 15, la Déclaration des droits de l'homme et du citoyen, préambule de la Constitution actuelle, précise : « *La société a le droit de demander compte à tout agent public de son administration* » (en 1789, cette disposition visait à se débarrasser des fermiers généraux). L'évaluation analyse des systèmes d'action,

pour comprendre la chaîne des causes, et savoir à quoi attribuer le résultat (imputabilité). Enfin, l'évaluation doit contribuer à éclairer la décision, stratégique ou opérationnelle.

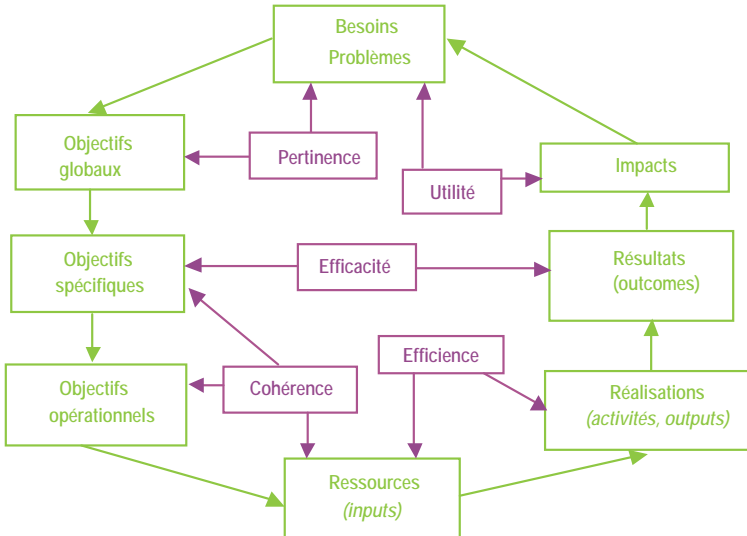
L'évaluation de l'action collective peut porter tout autant sur un projet que sur un dispositif de politique publique, un programme ou des politiques plus larges. Plus le degré de généralité est élevé, plus l'évaluation devient complexe. Pour autant, il est indispensable de procéder à ces évaluations larges, afin de repérer les effets de système. En effet, une politique publique se compose généralement de plusieurs programmes et plusieurs actions qui font parfois système et parfois se contredisent. Toute l'évaluation devra donc en tout premier lieu choisir le niveau auquel elle se situe. Il est souvent préférable de commencer par un niveau intermédiaire.

... qui comporte plusieurs registres...

Toute politique relève d'une théorie implicite du changement social. La mise en place d'une politique publique vise à résoudre les problèmes qui apparaissent. Par exemple, l'agitation dans les banlieues peut traduire un mécontentement face à un cadre de vie dégradé et à une problématique de logement. Face à l'agitation, le pouvoir politique peut décider de calmer les banlieues en réhabilitant le cadre de vie. Il s'agit là d'un objectif général, qui devra être décliné en programmes divers, répondant à des objectifs spécifiques, pouvant être eux-mêmes déclinés en objectifs opérationnels (schéma 1). Dans le cas de la sécurité routière, l'objectif général vise à diminuer le nombre des morts sur les routes. Plusieurs programmes spécifiques peuvent y contribuer, comme un programme d'aménagements routiers visant à supprimer les points noirs ; ou un programme visant à faire respecter le code de la route. Chacun des programmes spécifiques se décline à son tour en programmes opérationnels qui se déclinent en actions. Pour le premier, ce sera par exemple de supprimer tel passage à niveau ou de refaire tel carrefour. Concernant le second, ce pourra être de faire des contrôles de vitesse ou de refondre l'apprentissage de la conduite.

SCHEMA 1

Les registres de l'évaluation



Lecture • Les objectifs globaux sont définis à partir de l'identification des besoins ou problèmes ; l'adéquation des objectifs globaux aux besoins ou problèmes permet de mesurer la pertinence d'une politique.

À chaque politique, programme ou action, seront affectés des moyens spécifiques en termes humains, financiers et juridiques. Au niveau le plus bas – à savoir les objectifs opérationnels – sont mesurées les **réalisations** qui correspondent à l'activité de l'administration ou de l'opérateur dans un domaine particulier. C'est à ce niveau que l'on positionne les indicateurs d'activité, déclinés dans la loi organique relative aux lois de finances (LOLF). Pour reprendre l'exemple de la sécurité routière et de la réduction du nombre de points noirs, l'indicateur de réalisation pourra être le nombre d'appels d'offres réalisés par le service chargé des infrastructures ou la dépense engagée ; dans le cas de la lutte contre les vitesses excessives, ce sera le nombre de radars installés. Au niveau des objectifs spécifiques correspond l'évaluation des **résultats**. Ainsi, dans le domaine de la sécurité routière, le résultat sera le nombre de points noirs supprimés. Au niveau le plus élevé, celui des objectifs globaux, ce sont les **impacts** qui sont mesurés, généralement à plus long terme. La baisse du nombre des tués sur les routes constitue ainsi un impact de la politique de sécurité routière.

Parfois apparaissent des impacts inattendus. Par exemple, le fait de repeindre les cages d'escalier d'un quartier d'habitations à loyer modéré (HLM) a pu être à l'origine d'affrontements entre bandes de quartiers rivaux, dans la mesure où l'un des deux quartiers a vu son *standing* baisser par rapport à l'autre. Le simple fait de rénover un quartier a pu ici nuire à la tranquillité publique, contrairement à ce qui était espéré ; c'est un impact inattendu. Pour cerner au mieux les impacts d'une politique, il est donc important d'adopter une vision large avec ouverture d'esprit.

L'évaluation d'une politique publique se décline selon un certain nombre de critères, à commencer par sa **pertinence**, qui vérifie si l'action menée était en adéquation avec le problème qui était posé. Ensuite, il faut mesurer la **cohérence** des actions entre elles et avec les moyens déployés. Se posent aussi les questions de l'**efficacité** (les réalisations, les résultats et les impacts obtenus sont-ils conformes aux objectifs ?) ou encore celle de l'**efficience** (celle-ci rapporte les résultats obtenus aux dépenses engagées). Par exemple, l'efficience d'une politique d'emploi amène à mesurer le coût du chômeur évité ou de l'emploi créé, et à le comparer à celle d'un autre dispositif ou à celle d'un pays voisin. Enfin, le critère d'**utilité sociale** consiste à repérer si les impacts obtenus répondent effectivement aux besoins et aux problèmes d'aujourd'hui. Il s'agit là d'un critère fondamental en matière sociale, même s'il s'avère assez complexe de l'évaluer finement.

La rationalisation des choix budgétaires (RCB) s'appuyait sur une série d'indicateurs de moyens, d'activité et de résultats à court et long terme qui recouvrent les notions de ressources, réalisations, résultats et impacts. Le vocabulaire anglo-saxon est nettement plus clair : les moyens ou ressources affectés à l'action sont les « *inputs* » de l'activité qui produit des « *outputs* » (réalisations) qui engendrent des « *outcomes* » (résultats) à court terme et des « *impacts* » à long terme.

... et se distingue du contrôle, du pilotage et de l'audit

Il faut distinguer nettement l'évaluation, du contrôle, du pilotage et de l'audit (tableau 1). L'Inspection générale des affaires sociales (IGAS) et la Cour des comptes sont deux organismes chargés avant toute chose du contrôle. Le pilotage se développe au travers de la LOLF. L'audit connaît un succès parallèle, même s'il demeure encore peu répandu dans le monde administratif. À ce propos, la société française de l'évaluation (SFE) a organisé récemment une journée de travail en lien avec l'AFIGESE (associations des financiers et contrôleurs de gestion des collectivités territoriales) et « France qualité publique » pour travailler autour de ces notions. Leur travail a abouti à la publication d'un Cahier comparant ces différentes démarches : « Voisins : de la concurrence à la complémentarité ».

Ces démarches se distinguent de par leur objet, leurs normes de référence et leurs conséquences. Chacune correspond à un type de regard porté sur l'action publique :

- **Le contrôle** permet de vérifier la conformité par rapport à des textes (loi, réglementation, décret ou circulaire). Lorsqu'il y a un écart entre la pratique et la règle, la conséquence en est une sanction (amende, voire poursuites judiciaires).
- **Le pilotage** (contrôle de gestion) permet de suivre l'exécution des actions. La référence n'est plus la règle juridique ou comptable, mais le programme fixé en début d'année. Par exemple, chaque territoire s'est vu assigner un nombre de contrats à passer en matière de dispositifs d'aide à l'emploi. Chaque mois, est vérifiée la conformité du nombre de contrats effectivement passés avec l'objectif. En cas d'écart, se met en place un dialogue de gestion, qui permet de rectifier la trajectoire ou de déplacer des crédits.
- L'objectif principal **d'un audit** est la réduction des risques. L'auditeur (externe ou interne) se réfère pour cela aux standards de la profession; il ne s'agit pas de normes juridiques (comme dans le cas du contrôle) ou de programmes fixés à l'avance (comme dans le cas du pilotage ou du contrôle de gestion), mais du référentiel de la profession (bonnes pratiques préexistantes répertoriées); en cas d'écart, l'auditeur émet des alertes, des recommandations que l'opérateur peut, s'il le souhaite, ignorer.
- Enfin, **l'évaluation** a pour objectif d'améliorer l'action publique, d'en optimiser les résultats et pour cela de mesurer et d'expliquer les écarts entre les effets attendus et les effets obtenus. La norme en l'occurrence n'est autre que les objectifs de l'action. Elle se traduit dans la pratique par une série d'analyses qui permettent de vérifier si les effets correspondent à ce qui était attendu. Ainsi, elle est une aide à la décision opérationnelle et stratégique, et donne du sens à l'action.

TABLEAU 1

Contrôle, pilotage, audit et évaluation

	Contrôle	Pilotage (contrôle de gestion)	Audit	Évaluation
Objet	Vérifier la conformité à la réglementation	Suivre l'exécution des actions	Réduire les risques	Optimiser les effets-besoins Expliquer les écarts entre effets attendus et atteints
Normes, référentiels	Loi, règlement (cadre comptable)	Programme fixé <i>ex-ante</i>	Standards professionnels	Référence à construire en fonction des objectifs
Conséquences	Sanctionner un écart : amendes, poursuites judiciaires	Rectifier la trajectoire : dialogue de gestion	Alerter : recommandations, rappels à l'ordre	Aider à la décision (stratégique ou opérationnelle) selon les besoins de la société

Lecture • La démarche de pilotage analyse la conformité de l'exécution des actions au programme fixé *ex-ante*.

De façon simpliste et un peu caricaturale, on peut dire que le contrôle s'arrête à la procédure ; le pilotage s'intéresse d'abord aux réalisations et parfois intègre les résultats, alors que l'évaluation s'intéresse aux résultats et aux impacts. En ce sens, toutes ces démarches sont complémentaires.

L'évaluation répond à plusieurs objectifs et comporte plusieurs phases

L'évaluation peut répondre à quatre types et usages distincts et complémentaires. On distingue évaluation d'impact et évaluation de processus, d'une part ; aide à la décision et apprentissage collectif, d'autre part. L'évaluation d'impact vise à apprécier, de la manière la plus objective possible, les effets d'un programme sur la société. L'évaluation de processus se penche sur la mise en œuvre de cette politique et en analyse les processus : est ainsi analysé le système d'acteurs et la façon dont celui-ci a pu favoriser ou gêner l'atteinte des objectifs. S'agissant des acteurs de l'action publique, l'évaluation peut être une aide à la décision ; commandée et destinée aux décideurs ; et/ou utile à l'ensemble des acteurs dans un apprentissage collectif. Le processus d'évaluation permet aux agents de mieux comprendre les actions auxquelles ils participent et de s'en approprier les objectifs. Dans le cadre de certains dispositifs, ce type d'évaluation a permis de pointer l'existence d'actions inutiles.

Outre le choix du domaine qui doit être évalué et les décisions prises à l'issue de l'évaluation, deux prérogatives des décideurs (commanditaires), il est possible de distinguer quatre grandes étapes de l'évaluation.

La première a pour but de construire le référentiel de l'évaluation, en définissant les questions qui se posent et en retrouvant ou explicitant les objectifs initiaux. Parfois, cette première étape s'apparente à de l'archéologie administrative, quand les programmes publics se sont empilés successivement au cours du temps.

La deuxième étape consiste à recueillir les données. Généralement, le contrôle de gestion (pilotage), quand il existe, a effectué un suivi de réalisations, mais il va rarement aux résultats. Bien souvent, des enquêtes et des études *ad hoc* sont nécessaires. Des enquêtes existantes peuvent être réutilisées dans le cadre de l'évaluation, mais il est parfois nécessaire d'en entreprendre de nouvelles. Ces données et ces études permettent d'établir des constats.

Au cours de la troisième étape, on passe des constats aux conclusions. C'est à ce niveau qu'intervient l'interprétation. Il faut insister sur le fait qu'une étude n'est jamais à elle seule une évaluation : elle ne fait que participer à l'évaluation. Ainsi dans le cas des 35 heures, plusieurs études avaient été réalisées ; certaines disaient que cette politique avait créé des emplois, tandis que d'autres qu'elle en avait détruit. En fait, leurs résultats dépendaient des hypothèses retenues, éclairant chacune une facette de la réalité. Comme il n'existait aucun lieu qui permît de discuter les hypothèses sur lesquelles s'appuyaient ces études et d'en apprécier leur domaine de validité, cette apparente contradiction, non élucidée par les parties prenantes faute de lieu de débat apaisé, autorisait les luttes politiciennes.

Une étude donne un élément de constat, qui peut ensuite être interprété de différente manière selon le point de vue de chaque groupe d'acteur. Par exemple : le même chiffre sur le nombre de lits vacants dans un établissement d'hébergement pour personnes âgées dépendantes, va être interprété comme mauvais par le directeur financier qui y voit une perte de recettes potentielles ; comme bon par les agents de santé qui y voient un allègement de leur charge de travail, mais également par l' élu en charge de la politique de maintien à domicile des personnes âgées car cette vacance de lits permet l'hébergement d'urgence, corollaire de sa politique. Cette interprétation constitue la troisième étape de l'évaluation, avant la formulation des recommandations. Il est utile que cette phase puisse

se faire dans un cadre qui permette d'exprimer les désaccords et de « sortir la pépite du désaccord de la gangue des malentendus ».

De là on passe (ou non) à la phase des recommandations, qu'il est utile de tester ou de coproduire avec les principaux intéressés, ceux qui devront les appliquer. D'où l'intérêt des comités de pilotage de l'évaluation ou instances regroupant les parties prenantes, pour faciliter son appropriation par les acteurs, tant en termes d'appréciations et d'interprétations qu'en termes de suivis opérationnels.

Les aspects historiques de l'évaluation en France : faiblesses et aléas

Des débuts difficiles au niveau central

Un décret de 1990 crée le Comité interministériel de l'évaluation assorti d'un Conseil scientifique, chargé de définir les concepts et méthodes de base d'un champ qui restait fort peu exploré en France. Il lui a été reproché sa lenteur et son activité limitée, ce qui s'expliquait à la fois par le fait que l'évaluation en France était en phase d'apprentissage, mais aussi et surtout par son positionnement interministériel, lieu de la complexité la plus grande des politiques étudiées et de l'absence de responsabilité de chacun des ministres. En 1998, les deux instances sont fusionnées en un Conseil national de l'évaluation (CNE) ; l'expérience accumulée lui a permis une plus grande efficacité : il a produit en quatre années d'existence autant d'évaluations qu'en avait produites le CIME en huit ans. Mais les difficultés du positionnement interministériel demeuraient. Par ailleurs, l'évaluation était positionnée du côté de l'exécutif seulement, et non du législatif, encore moins en surplomb des deux.

Entre 2000 et 2006, aucune instance nationale n'a été chargée de l'évaluation des politiques publiques. Côté législatif, le Parlement avait mis en place une mission d'évaluation et de contrôle, commune aux deux assemblées, Assemblée nationale et Sénat, qui n'a jamais fonctionné ; l'Assemblée nationale a alors créé une mission d'évaluation et de contrôle autonome (MEC) procédant par auditions et utilisant les enquêtes réalisées par l'administration. Cependant, cette institution manquait notablement de moyens. En 2006, avec la LOLF, apparaissait la notion de « performance » (anglicisme pour « réalisations »), première étape vers un retour de l'évaluation. Le secrétariat d'État à la prospective et à l'évaluation des politiques publiques, créé en 2007, a disparu en janvier 2009, et la petite Mission d'évaluation et prospective qu'il avait constituée au Centre d'analyse stratégique s'est réfugiée auprès du ministère du Budget.

Des perspectives intéressantes

La révision constitutionnelle de juillet 2008 offre des perspectives intéressantes, dans la mesure où elle inscrit la notion d'évaluation des politiques publiques au sein même de la Constitution. La LOLF avait tracé la voie d'une évaluation de l'action du pouvoir exécutif par le législatif, mais le Parlement ne s'était pas jusqu'alors véritablement saisi de cette capacité. Suite à la réforme constitutionnelle, l'Assemblée nationale a créé en juin 2009 un Comité d'évaluation et de contrôle des politiques publiques (remplaçant la MEC), coprésidé par un membre de la majorité et de l'opposition, chargé de définir un programme d'évaluation.

La Cour des comptes, qui peut venir en appui au Parlement, pourrait l'aider dans cette tâche. Cependant, elle est composée par définition de magistrats, amenés à porter des jugements et dont la culture est celle du contrôle. La gestion de la pluralité des points de vue ne lui est pas

familière. Cependant, la Cour met en place une réforme des institutions financières de contrôle (nationale et régionale) pour mieux répondre à ses nouvelles missions (certification des comptes de l'État avec la LOLF, et maintien de l'évaluation avec la réforme constitutionnelle) en distinguant quatre métiers qui sont quatre regards sur l'action publique : contrôleur, juge, certificateur, évaluateur.

Les collectivités territoriales sont en avance

Les collectivités territoriales n'ont pas suivi le même chemin que l'État. En effet, au moment où le principe de l'évaluation perdait du terrain au niveau central, il connaissait un véritable essor au niveau territorial. En ce sens, les administrations centrales feraient bien de s'inspirer du niveau décentralisé. Cet écart s'explique en partie par les contraintes d'évaluation imposées aux collectivités territoriales par les politiques contractuelles (Fonds européens, contrats de plan État-régions ou les contrats de la politique de la ville). Les Conseils régionaux, les Conseils généraux les plus importants ont mis en place des fonctions d'évaluation ; au début elles sont le plus souvent rattachées aux services de contrôle de gestion et d'audit ; puis prenant de la distance, elles sont placées directement auprès de la Direction générale des services ou du président. Ces nombreuses expériences de terrain sont propices à l'innovation. Le lien entre le suivi des mesures et l'évaluation (suivi évaluatif) donne lieu à un développement inédit de nouvelles méthodes.

Dans le cadre de l'évaluation des fonds européens, et en conformité avec les standards d'évaluation prônés par la Commission européenne, la France a mis en place des instances d'évaluation qui rassemblent les acteurs concernés : l'Instance nationale d'évaluation de programmes opérationnels du FEDER (Fonds européen de développement régional) qui est coprésidé par un universitaire et un président de Conseil Régional, et dont le secrétariat est assuré par la DIACT (Délégation interministérielle à l'aménagement et à la compétitivité des territoires) ; c'est également le cas pour le FEADER (Fonds européen agricole pour le développement rural) géré par le ministère de l'Agriculture. Le Conseil scientifique d'évaluation du Fonds social européen (FSE), que je préside, a également vu le jour l'an dernier.

Le rôle de la Société française de l'évaluation

En conclusion, je tiens à faire un bref retour sur la mission de la SFE. La Société française de l'évaluation, association loi 1901, a pour objet de promouvoir la culture de l'évaluation et sa qualité. Elle rassemble toutes les catégories d'acteurs concernés par l'évaluation (commanditaires, universitaires, chercheurs, prestataires publics ou privés). Pour améliorer la qualité des évaluations et leur bon déroulement, elle a mis en place un groupe travaillant sur les standards et la déontologie qui a rédigé une charte de l'évaluation. Cette charte est mise en ligne sur le site de la SFE (www.sfe-asso.fr). Parmi les grands principes qu'elle porte, la charte insiste sur la pluralité des points de vue, sur la compétence des évaluateurs, sur la distanciation (indépendance de l'évaluateur). Le respect des personnes, notion chère aux statisticiens, la transparence (de la méthode, des résultats) y figurent également. À propos de cette dernière notion, il faut souligner l'importance de la publication des résultats qui doit être décidée (ou non) avant de les connaître ! Enfin le principe d'opportunité rappelle qu'il ne sert à rien de lancer une évaluation dès lors que l'on sait qu'elle n'aura aucune utilité.

Cette charte a pour objectif premier de réguler le marché entre les commanditaires et les prestataires, qu'ils soient privés, universitaires, chercheurs ou statisticiens. Un nombre tou-

jours plus important de commanditaires (comme la DIV¹, l'ANRU²...) inclut la charte en annexe dans leurs appels d'offres, ce qui permet aux prestataires privés de faire valoir certains principes importants et d'éviter certaines dérives (éviter la tentation d'acheter les résultats d'une étude avant même qu'elle ne soit lancée). La SFE organise régulièrement des journées sur des thèmes variés (évaluation de l'activité et de la qualité des prestations des établissements socio et médico-sociaux dans le cadre de la loi de 2002 ; l'évaluation de l'utilité sociale de l'action publique ; évaluation et statistique, évaluation et communication, etc.). Créée avec l'appui du Centre d'études et de recherche sur les qualifications (CEREQ) et du Centre d'études de l'emploi, la SFE a fêté ses 10 ans à Marseille, les 11 et 12 juin 2009, en présence de Philippe Séguin, premier président de la Cour des comptes et de Didier Migaud, président de la Commission des finances de l'Assemblée nationale³.

L'évaluation constitue une aide à la décision stratégique et opérationnelle, dans un processus d'apprentissage collectif qui donne du sens à l'action et favorise le changement, en vue d'une amélioration continue de l'action publique.

1. Délégation interministérielle à la ville (DIV).

2. Agence nationale de rénovation urbaine (ANRU).

3. Devenu depuis premier président de la Cour des comptes, succédant à Philippe Séguin. Voir l'ouvrage tiré de ces Journées, Mouterde, Trosa, 2010.

L'évaluation d'impact en économie du développement

L'approche expérimentale

Esther Duflou (Collège de France)

En 1932, au beau milieu d'une période qui ressemble étrangement à celle que nous vivons aujourd'hui, Franklin Delano Roosevelt déclarait : « *Notre pays a besoin de – et à moins que je ne me méprenne sur ses aspirations – notre pays exige que nous essayions de nouvelles choses, avec audace et persévérance. Tenter une nouvelle approche relève du bon sens : si elle échoue, admettons-le avec franchise et essayons-en une autre. Mais, avant toute chose, tentons quelque chose.* » C'est dans ce cadre que se place le travail d'évaluation. Il ne doit pas être compris de manière restrictive, comme une sanction après coup. Il doit aider à la prise de décision et à l'expérimentation politique.

La lutte contre la pauvreté n'échappe pas à cette approche. Il faut sans cesse expérimenter et réfléchir en dehors des cadres habituels. Malheureusement, le domaine de la politique sociale manque trop souvent d'imagination. Il faut y voir en partie l'effet de contraintes politiques. Les hommes politiques sont bien souvent prisonniers de leurs ambitions. Ils voudront toujours en faire trop et en finir avec un problème sur-le-champ. Les politiques sont souvent mises en place au niveau national, avant même d'avoir été testées sur des territoires plus restreints. C'est pourtant ainsi que les idées peuvent être confrontées à la réalité du terrain et par la suite affinées.

Les chercheurs, qu'ils dépendent de l'université ou des administrations, en collaborant avec les partenaires de terrain (État, collectivités locales, organisations non gouvernementales, ONG) peuvent apporter une contribution utile dans ce processus d'expérimentation créative. D'une part, ils peuvent identifier ou proposer des solutions nouvelles. D'autre part, ils savent évaluer rigoureusement les politiques publiques. C'est cette articulation entre prospective et évaluation qui a connu un fantastique essor au cours des dix dernières années dans le champ de l'économie du développement.

L'évaluation d'impact

Économistes, administrations et hommes politiques peuvent se tromper dans leurs recommandations et suggestions. En effet, toute recommandation s'appuie sur une analyse du réel, qui le simplifie inévitablement. Au cours de cette simplification, des aspects qui peuvent au final

s'avérer essentiels sont souvent perdus. Toute politique, comme le soulignait Annie Fouquet dans son intervention, peut avoir des effets inattendus, qui peuvent s'avérer finalement totalement contre-productifs. Pour autant, il serait regrettable d'en conclure qu'il est nécessaire pour les chercheurs de rester inactifs au prétexte qu'ils ne veulent pas être complices d'une erreur. Au contraire, étant donné que chacun peut commettre des erreurs, il est préférable de continuer à travailler avec les partenaires de terrain, en adoptant une méthode d'évaluation rigoureuse. C'est dans l'évaluation plutôt que dans l'inaction que réside le principe de prudence.

Pourquoi y a-t-il encore si peu d'études d'impact ?

Ces éléments semblent relever du bon sens. Cependant, dans les pays pauvres comme dans les pays riches, les études d'impact demeurent peu nombreuses. Comment expliquer ce constat ? Une première hypothèse consiste à affirmer que personne ne souhaite véritablement que soit pratiquée une évaluation rigoureuse. Les partisans d'un programme – hommes politiques qui souhaitent être réélus, fonctionnaires qui souhaitent être promus, ONG qui souhaitent continuer à percevoir un financement, multinationales qui souhaitent montrer qu'elles agissent de manière responsable pour la planète – ont tout intérêt à proclamer que leur action a été couronnée de succès. Dans le cas de la mise en place des 35 heures, il est évident que certains acteurs politiques préféreraient brandir certains résultats, tandis que leurs adversaires faisaient valoir d'autres études. En l'absence d'évaluation rigoureuse, il est toujours plus simple de présenter un programme comme un succès, même si personne n'est dupe, pas plus l'administration que les électeurs. L'habitude entraîne par ailleurs les acteurs dans une logique de surenchère permanente. Ainsi, si par souci de franchise, un homme politique, qui a défendu la mise au point d'un programme, convient que celui-ci a donné des résultats plutôt satisfaisants, au lieu de les présenter comme excellents, il risque de provoquer une certaine défiance des électeurs. Ceux-ci vont penser que cette politique n'a en réalité pas du tout donné les résultats escomptés. Des succès relatifs sont toujours suspects au final.

Cette perspective fort cynique est bien souvent exagérée. Beaucoup d'acteurs, au sein des administrations, des gouvernements ou des ONG, sont pleinement ouverts à la possibilité d'une évaluation, pour tirer les leçons des expériences passées et améliorer les dispositifs. En réalité, si l'évaluation demeure rare, c'est parce qu'elle est difficile à mener.

Comment expliquer la nécessité d'évaluer des politiques publiques, alors que cette exigence ne s'applique pas au secteur privé ? En effet, Renault n'a pas besoin d'évaluer ses véhicules mis sur le marché, alors que le gouvernement doit évaluer ses politiques sociales. Sur le marché privé, tout est plus simple, les consommateurs mécontents peuvent se tourner vers une autre marque. Par exemple, si les voitures Renault ne correspondent pas aux besoins des consommateurs, ces derniers vont acheter par exemple un modèle Peugeot. L'évaluation est donc automatique. Elle correspond au test du marché, alors que, dans le secteur public, il n'existe pas de choix.

Pourquoi évaluer l'impact d'une politique ?

En outre, il faut être conscient du fait que ces politiques sociales interviennent justement lorsque le marché s'avère défaillant pour répondre à un objectif que la société s'est fixée. Nous estimons ainsi qu'il est important que tous les enfants bénéficient d'une éducation. Par conséquent, nous sommes prêts à agir de manière commune, même lorsque les parents ne souhaitent pas ou ne sont pas en mesure de financer leurs études. En France, le choix a été fait de construire des écoles publiques. Au Chili, au contraire, l'État distribue des bons qui

permettent aux enfants de suivre une scolarité dans une école privée. Dans les deux cas, le gouvernement finance l'éducation.

Comment savoir laquelle de ces deux politiques s'avère la plus efficace? Comment savoir s'il est préférable de construire une multitude de petites écoles ou un nombre restreint de grands établissements? Comment savoir combien d'enfants mettre par classe? Comment savoir s'il est nécessaire de construire des groupes de niveau? Ces choix ont tous des conséquences sur la qualité de l'éducation. Les bénéficiaires de ces politiques sont contraints de prendre ce qu'on leur donne, à moins d'estimer que le système mis en place est tellement mauvais qu'il est préférable de suivre une scolarité privée, ainsi qu'on le constate en Inde, où même les familles les plus pauvres renoncent à envoyer leurs enfants à l'école publique, tant son niveau est faible. Lorsque l'on arrive à de telles extrémités, il est déjà trop tard. Il aurait fallu imaginer différents modèles, avant de pouvoir s'accorder sur la solution la plus efficace.

Les deux étapes de l'évaluation : évaluation de processus et évaluation d'impact

Pour s'assurer de l'impact d'une politique, il est donc nécessaire de procéder à une évaluation. Cette évaluation se décline en plusieurs étapes. Tout d'abord, l'évaluation du processus (audit, contrôle...) permet de vérifier que la politique a été mise en place comme prévu. Ensuite, une évaluation d'impact doit être menée. Elle permet de vérifier si les résultats escomptés ont été effectivement atteints.

La première de ces deux évaluations n'est conceptuellement pas difficile à réaliser, même si elle peut s'avérer délicate en pratique. En principe, il suffit de suivre les flux financiers et les réalisations effectives, afin de documenter ce qui a été entrepris et comparer cela au programme défini à l'avance.

L'évaluation d'impact est en revanche complexe en ce sens qu'elle impose de comparer la situation des bénéficiaires d'une politique à la situation qu'ils auraient connue si cette politique n'avait pas été mise en œuvre. Or ce dernier aspect relève de la spéculation. Il est possible de déterminer avec plus ou moins de précision la situation d'un bénéficiaire avant la mise en place de la politique et après celle-ci. Il serait cependant une erreur de croire que la situation d'un bénéficiaire avant la mise en place de la politique correspond à la situation que ce même bénéficiaire aurait connue si cette politique n'avait pas été mise en place. Beaucoup de choses auraient pu influencer entre-temps.

Les différentes méthodes d'évaluation d'impact

Pour résoudre ce dilemme, il est important de constituer un groupe de contrôle adéquat. Ce groupe test sera bénéficiaire d'une politique différente ou exclu de toute politique et totalement comparable dans sa composition au groupe des bénéficiaires. En réalité, la difficulté réside dans le fait que les non-bénéficiaires ne sont généralement pas vraiment comparables aux bénéficiaires. Ces derniers sont choisis en effet en fonction de critères précis qui les distinguent du reste de la population.

Prenons par exemple la question de la parité hommes-femmes. Aujourd'hui, près de cent pays de par le monde garantissent la représentation politique des femmes, parmi lesquels la France. Sur ce sujet, au cours de l'évaluation, il faut d'abord déterminer si les politiques mises en place conduisent à l'élection d'un plus grand nombre de femmes. Ensuite, il faut vérifier si la présence accrue des femmes a un impact sur le type d'actions entreprises et mesurer enfin la perception du public à l'égard des femmes dans leur rôle de décideur. Or il faut être conscient que les villes et les villages qui élisent une femme maire sont certai-

nement différents des autres. Ils ont vraisemblablement une meilleure image des femmes et des préférences politiques différentes. Il est possible d'imaginer que les électeurs prêts à élire une femme maire soutiennent par exemple les politiques d'aide à l'enfance.

Économétrie traditionnelle

Comment constituer un groupe test viable ? L'économétrie traditionnelle a recours à des techniques statistiques (*matching*, régression...) en contrôlant l'ensemble des variables, pour tenter d'effacer les biais *ex post*. Dans le cas de la parité, il serait ainsi possible de vérifier le poids relatif des femmes dans la population de la commune, l'éducation des femmes, les politiques entreprises par le passé... Cependant, on ne peut jamais être sûr d'avoir pris en compte tous les paramètres nécessaires. La manière dont l'on construit le groupe de contrôle peut avoir un impact très fort sur les résultats. En appliquant plusieurs méthodes de constitution du groupe de contrôle, il est possible au final de choisir les résultats qui arrangent le décisionnaire.

Expériences naturelles

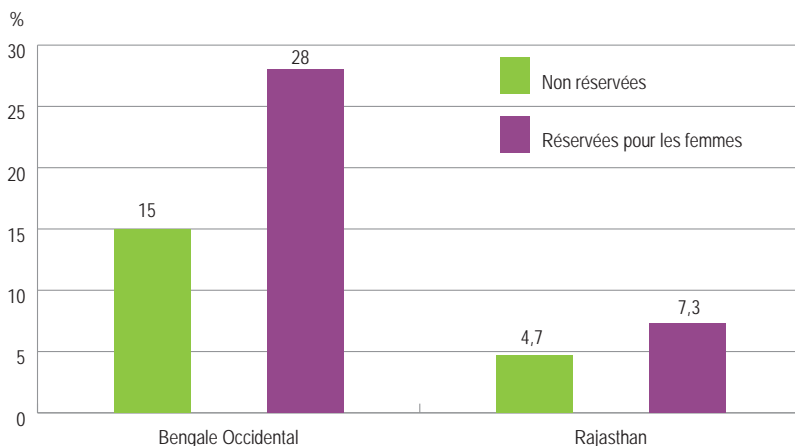
La deuxième possibilité réside dans les expériences naturelles, qui exploitent des situations dans lesquelles les groupes sont parfaitement comparables. Pour reprendre l'exemple de la parité, il est possible de comparer une commune où une femme a gagné de justesse à une autre où une femme a perdu de justesse. On peut estimer que c'est le hasard de l'élection qui a fait en sorte que certains partisans d'un candidat ne sont pas déplacés. Cependant, il n'est pas toujours possible de mettre en place ces expériences naturelles.

Évaluation aléatoire

La troisième méthode a pour nom « l'évaluation aléatoire ». Celle-ci est prospective. Avant même la mise en place du programme, est choisi aléatoirement parmi un groupe de bénéficiaires potentiels un échantillon de participants, qui deviendra le groupe test. Il faut réunir certaines conditions pour mettre en place une telle évaluation. Elle n'est possible que pour les programmes pilotes qui n'ont par définition pas vocation à prendre en charge l'intégralité d'une population ou lorsque les budgets sont trop limités pour traiter toute la population. Ce fut le cas pour la prise en charge de certains chômeurs de longue durée par des entreprises privées en France. Les programmes en phase d'expansion sont également propices à cette technique. En effet, il est possible de déterminer au hasard l'ordre dans lequel les programmes sont attribués. Enfin, parfois, il arrive qu'allouer un programme au hasard soit la solution retenue comme la plus équitable. Dans tous les cas, il est nécessaire de choisir le groupe test avant la mise en place de la politique. Le groupe de contrôle est alors parfaitement comparable au groupe des bénéficiaires. Parfois, ce sont deux politiques différentes qui sont appliquées à deux groupes identiques.

GRAPHIQUE 1

Investissement en eau potable dans deux États indiens



Source • Chattopadhyay et Duflo 2004.

Lecture • Dans les villages du Bengale occidental où la présidence du conseil municipal est réservée aux femmes, les investissements en eau potable représentent 28 % des dépenses totales de la municipalité.

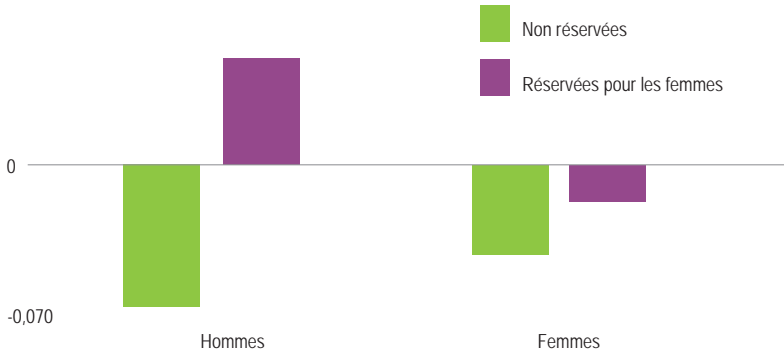
Dans le cas de l'évaluation de la politique de parité hommes-femmes en Inde, la « randomisation » a été mise en place pour des raisons politiques et non pour faciliter le travail. En effet, en 1993, un amendement à la constitution prévoit que les conseils municipaux et les conseils de villages devront désormais inclure un tiers de membres féminins et qu'un tiers des maires de groupes de villages (groupements de communes) devront être des femmes. Pour éviter les manipulations, les villages qui doivent élire une femme maire sont choisis au hasard. Dans ces villages, tous les candidats doivent être des femmes. Les décisions prises par des conseils municipaux dirigés par une femme ont été comparées à celles dirigées par des hommes. Il est ainsi apparu que les femmes privilégient les investissements dans l'accès à l'eau potable et relatifs aux besoins spécifiques des femmes (graphique 1).

Les discriminations à l'égard des femmes

Les évaluateurs se sont également intéressés à la discrimination à l'encontre des femmes. Il n'est pas simple d'évaluer ce critère, car les personnes interrogées ont tendance à présenter l'image qu'elles aiment avoir d'elles-mêmes. Généralement, dans les pays riches, elles sous-estiment leurs propres discriminations. En Inde, il s'agirait plutôt du phénomène inverse. Pour éviter ce biais, les chercheurs font écouter un même discours enregistré par un homme et par une femme. Les personnes interrogées n'écoutent qu'un seul des deux discours. Il apparaît que, dans les villages qui n'ont jamais eu de femme maire, la réaction est plus négative lorsque des hommes entendent le discours prononcé par une voix de femme, ce qui démontre l'existence d'une discrimination (graphique 2). En revanche, dans les lieux où une femme a été élue maire, le biais s'inverse. Les hommes ont ainsi tendance à plus apprécier le discours de la femme que celui de l'homme. Chez les femmes, les écarts sont plus faibles, avec un léger *a priori* positif pour les discours des hommes dans les deux cas.

GRAPHIQUE 2

Différence entre l'évaluation d'un discours prononcé par une femme ou par un homme au Rajasthan et au Bengale occidental



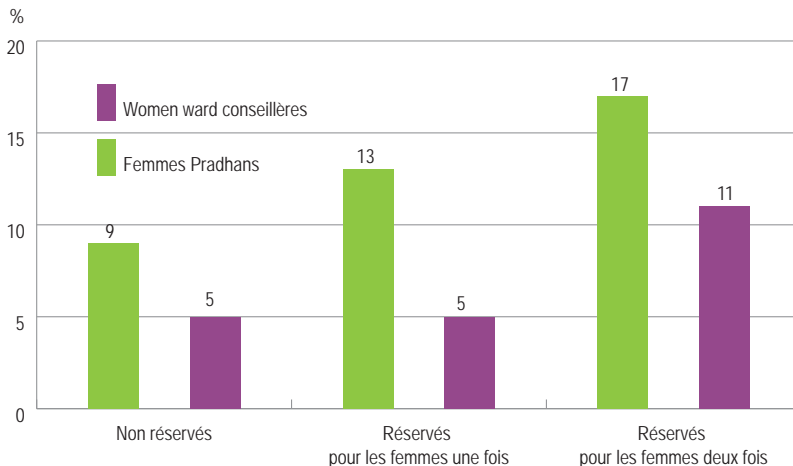
Source • Beaman, Chattopadhyay, Duflo, Pande et Topalova, 2008.

Lecture • En moyenne, l'évaluation par les hommes interrogés d'un discours prononcé par une femme sur une échelle de 1 à 10 est inférieure de 0,07 point à l'évaluation du même discours prononcé par un homme dans les villages non réservés à une femme alors qu'elle est supérieure dans les villages réservés à une femme.

Lorsqu'une commune a été réservée aux femmes pour un mandat, elle devient systématiquement non réservée lors du mandat suivant. Il s'avère que dans ces villages, les femmes maires sont plus nombreuses que dans les villages non réservés (17% pour les communes réservées deux fois, 13% pour les communes réservées une fois, contre 9% pour les communes qui n'ont jamais été réservées, voir graphique 3). Ces chiffres démontrent l'existence d'un impact de cette politique sur l'élection de femmes, du fait de la réduction des discriminations.

GRAPHIQUE 3

Pourcentage de femmes élues sur des sièges non réservés au Rajasthan et au Bengale occidental



Source • Beaman, Chattopadhyay, Duflo, Pande et Topalova, 2008.

Lecture • En moyenne, dans les villages réservés aux femmes une fois par le passé, 13% des maires et 5% des conseillers municipaux sont des femmes contre respectivement 9% et 5% dans les villages où le poste de maire n'a jamais été réservé aux femmes.

L'évaluation expérimentale

L'évaluation aléatoire s'est développée aux États-Unis dans les années soixante. Elle était pratiquée essentiellement pour les grands programmes, relatifs notamment à la réforme de l'État-providence. Ainsi, ont pu être évalués des programmes de formation professionnelle ou de responsabilisation. Il y a dix ans, cette notion restait quasi inconnue dans les pays en développement. Aujourd'hui le réseau de chercheurs spécialisés dans la lutte contre la pauvreté Jameel Poverty Action Lab travaille sur une centaine de projets d'évaluation dans les pays en voie de développement ou dans les pays pauvres. Par ailleurs, la Banque mondiale, l'Agence française de développement (AFD), la *Millenium challenge corporation (MCC)*, de même que le Haut commissariat aux solidarités actives contre la pauvreté financent des évaluations aléatoires.

La première génération de ces évaluations essayait avant toute chose de déterminer quelles actions donnent des résultats positifs. De la sorte, les chercheurs pouvaient voir si les opérations de déparasitage réduisaient l'absentéisme scolaire, si la réduction de la taille des classes ou l'envoi de livres jouait de manière bénéfique sur les résultats. Aujourd'hui, ces évaluations prennent une autre forme. Elles sont souvent perçues comme un outil permettant de tester des théories. Dans cette approche plus souple, l'évaluation n'est pas véritablement utilisée comme un outil de bilan, mais plutôt comme un outil de prospective. Les partenaires de terrain voient en même temps si leur programme fonctionne et comment l'améliorer, car ce sont en général plusieurs versions du même programme qui sont testées simultanément. Grâce à ce cadre théorique, les expériences se renforcent mutuellement. Chaque évaluation participe de fait à un processus d'accumulation des expériences, au profit de la connaissance collective.

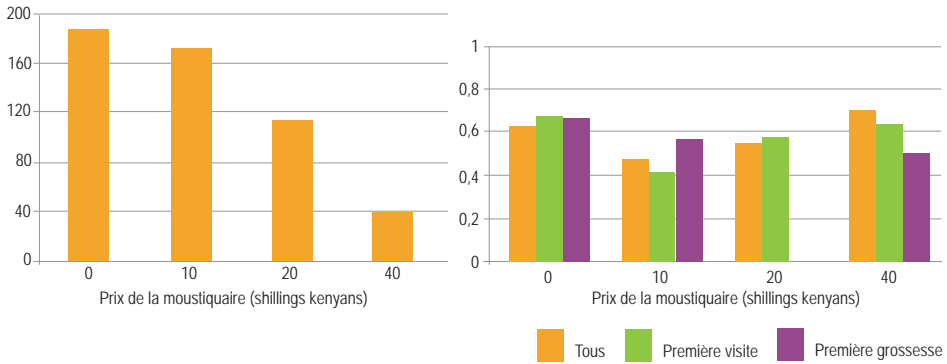
La gratuité comme solution au paludisme

Par exemple, la distribution de moustiquaires imprégnées d'un produit insecticide est considérée comme la meilleure méthode de lutte contre le paludisme. Ces produits coûtent en moyenne 5 euros à l'unité. La question se pose sur l'opportunité de distribuer ces produits gratuitement. En faveur de la gratuité, certains font valoir des effets externes positifs (il s'avère que les moustiques tendent à fuir les zones où une large partie de la population a été équipée) et une élasticité prix importante. De plus, les populations sont, selon eux, découragées par des prix mêmes faibles. À l'encontre de cette vision, d'autres exposent un argument relatif à l'effet de sélection. Les personnes qui n'ont pas besoin de moustiquaire se voient distribuer un produit qu'elles ne vont pas utiliser. Parfois, les moustiquaires sont recyclées pour faire des rideaux. Par conséquent s'instaure un certain gaspillage. De plus, le fait de payer pour un objet le rend soudainement plus précieux. Donc ceux qui payent l'utilisent plus fréquemment. Le nombre de moustiquaires distribué est ainsi plus faible, mais le taux d'utilisation peut s'avérer plus important.

Jessica Cohen et Pascaline Dupas ont testé ces deux hypothèses, en faisant varier de façon aléatoire le prix des moustiquaires proposées dans les maternités au Kenya. L'argument de l'élasticité-prix élevé se vérifie effectivement. Lorsque la moustiquaire est gratuite, ce sont 190 unités en moyenne qui sont distribuées chaque mois, alors que ce chiffre tombe à 40 lorsque le prix est fixé à 40 shillings, soit moins d'un dollar, le sixième du prix du produit. Il apparaît en revanche que les moustiquaires sont tout autant utilisées lorsqu'elles sont distribuées gratuitement que lorsqu'elles le sont contre 40 shillings. Cette étude a été conduite au moment où un débat virulent s'était engagé entre l'Organisation mondiale de la santé d'une part, qui prônait une distribution gratuite et le mouvement du marketing social, d'autre part, partisan de la distribution payante (graphique 4).

GRAPHIQUE 4

Nombre de moustiquaires distribuées (à gauche) et pourcentage de ménages utilisant les moustiquaires (à droite)



Source • Cohen et Dupas, 2009.

Lecture • 170 moustiquaires coûtant 10 KSh ont été distribuées et ces moustiquaires sont utilisées dans 48 % des cas.

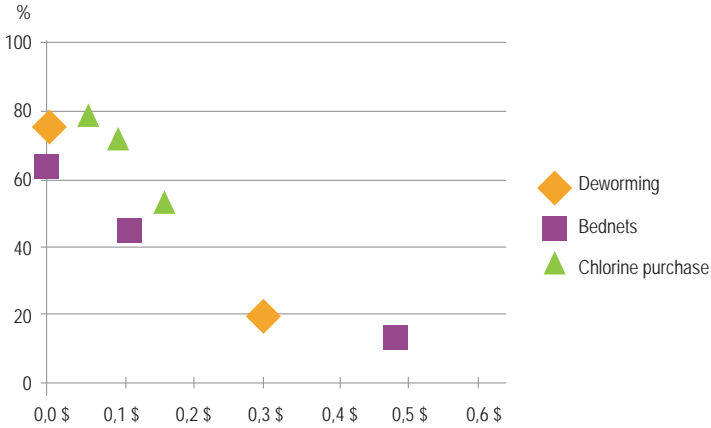
Dani Rodrik publia alors les résultats de cette étude sur un blog très lu sous le titre « Jeff Sachs a raison », Jeff Sachs étant un économiste proche de l'OMS. Sur ce même blog, un autre économiste, Mead Over, protesta, avançant l'argument que ces résultats ne valaient que pour les femmes enceintes et faisant valoir que le choix du Kenya était biaisé, dans la mesure où le marketing social y est actif depuis longtemps. Dani Rodrik a alors réagi en retirant ses premières conclusions. Parallèlement le débat a progressé avec cette étude puisque s'ouvre la perspective d'une expérimentation sur d'autres populations que les femmes enceintes kenyanes. De plus, *Population services international* (PSI), qui se chargeait de la distribution, a à ce moment décidé de distribuer gratuitement les moustiquaires dans les maternités du Kenya. Pascaline Dupas a donc renouvelé l'expérience au Kenya, en distribuant des bons sur les marchés à des femmes et des hommes. Les résultats obtenus sont identiques. Deux autres chercheuses ont développé une expérience similaire à Madagascar et en Ouganda, pour des résultats allant dans le même sens.

La demande de produits de santé est très sensible au prix

Les résultats de l'étude sur les moustiquaires ont fait apparaître une élasticité-prix forte. Reste maintenant à vérifier si cette élasticité se retrouve pour d'autres produits. Des expériences ont été réalisées sur de nombreux soins préventifs. Deux types de conclusions en ressortent. D'une part, les prix, même faibles, découragent les usagers. Cela se vérifie pour le déparasitage au Kenya, où le fait de faire payer 30 shillings a abouti à une réduction de la demande de 75 %. Une étude menée sur les désinfectants pour l'eau, aussi bien au Kenya qu'en Zambie, semble aller dans le même sens (graphique 5).

GRAPHIQUE 5

Proportion de personnes achetant un traitement en fonction de son prix dans le district de Busia (Kenya)



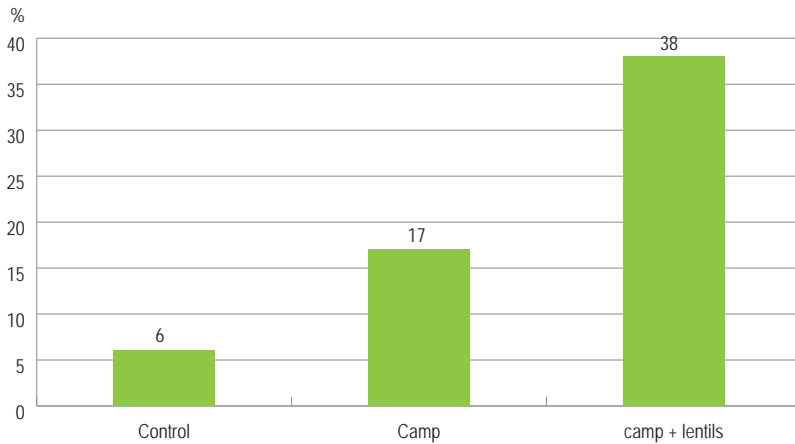
Lecture • 70% des personnes achètent un traitement à base de chlorure coûtant 0,10 \$ (Deworming = déparasitage, Bednets = moustiquair, Chlorine purchase = Traitement à base de chlorure).

Et à de légères incitations

Les prix négatifs (légère incitation financière) semblent au contraire fortement inciter les populations, comme l'a notamment montré une étude sur la vaccination dans le district d'Udaipur au Rajasthan qui a confirmé cette dernière conclusion. Les taux de vaccination y atteignent 5%, alors que la vaccination est censée être gratuite. Dans les faits, le manque d'infirmières et l'éloignement des centres de vaccination réduit l'offre, alors que les populations ne sont pas véritablement sensibilisées à ces problèmes, ce qui réduit aussi la demande. Dans 60 villages, une ONG a mis en place une campagne de vaccination. Un infirmier se rend dans chacun d'entre eux tous les mois en moto pour y vacciner les enfants. Le problème de l'offre se trouve ainsi résolu. Une incitation a été mise en place, puisque les personnes qui font vacciner leur enfant se voient remettre un kilo de lentilles. Ce cadeau est bien entendu trop faible pour convaincre quelqu'un qui est opposé à la vaccination. Là où ces camps de vaccination itinérants ont été mis en place, le taux de vaccination a atteint 17%. Il atteint même 38% avec la distribution d'un kilo de lentilles. Ainsi, le coût de la vaccination se réduit de manière notable avec la distribution de lentilles (graphique 6).

GRAPHIQUE 6

Proportion d'enfants ayant reçu au moins deux vaccinations dans certains villages du district d'Udaipur en Inde



Source • Bannerjee, Duflo, Glennester et Kothari, 2010.

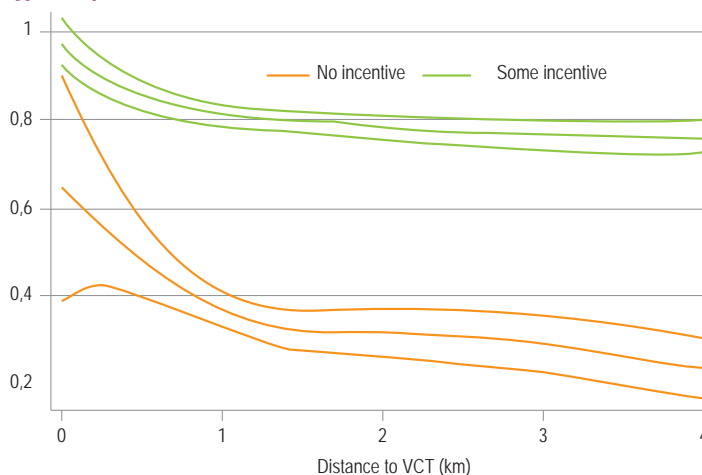
Lecture • Dans les villages concernés par la campagne de vaccination mais où aucune incitation n'a été mise en place, 17% des enfants ont reçu au moins deux vaccinations.

Rebecca Thorntorn, chercheuse à l'Université du Michigan, a mené une étude de même nature sur le dépistage du SIDA au Malawi. Le conseil et dépistage volontaire (VCT) est en effet considéré en Afrique comme la pierre angulaire de la lutte contre la maladie. Au Mozambique, par exemple, cette politique représente 55% des fonds consacrés à ce problème. Pour autant, très peu de personnes subissent un dépistage. Souvent, l'on considère que des barrières psychologiques ou sociales dissuadent les populations d'effectuer un dépistage. Dans les pays où les antirétroviraux ne sont pas disponibles, le fait de savoir que l'on est séropositif équivaut à une sentence de mort. Socialement, le fait d'être porteur du virus peut entraîner des discriminations fortes.

Dans le cadre de l'enquête menée par Rebecca Thorntorn, 2 812 personnes ont effectué un test de dépistage. Ces personnes étaient invitées à venir chercher, si elles le souhaitaient, leurs résultats quelques semaines plus tard. À la fin de l'interview, le participant se voyait ou non remettre une capsule de bouteille sur laquelle était inscrit un numéro, correspondant à un cadeau s'il venait chercher ses résultats. Parallèlement, l'emplacement de la tente où les personnes pouvaient aller chercher leurs résultats était choisi aléatoirement, soit dans un endroit proche du village, soit plus loin. Le taux de personnes venant chercher leurs résultats passe de 35 à 65% avec une incitation très faible (proche de 10 cents). Ceci démontre que les barrières psychologiques ne sont pas aussi importantes que ce qui est annoncé et que l'incitation financière est forte. Par ailleurs, plus le centre est loin et plus le taux de personnes venant rechercher leurs résultats faiblit (graphique 7).

GRAPHIQUE 7

Pourcentage de prise de connaissance des résultats



Source • Thornton 2008.

Lecture • Dans les villages où aucune incitation n'a été mise en place, on estime que 37 % des personnes habitant à 1 kilomètre du centre de dépistage sont allées chercher les résultats contre 82 % dans les villages où une incitation a été mise en place. Les vraies valeurs des paramètres ont 95 % de chances de se trouver respectivement dans les intervalles [34 %, 40 %] et [80 %, 84 %].

Un effet d'apprentissage

Suite aux études menées dans plusieurs pays sur la distribution des moustiquaires est apparu un nouvel argument. Certains prétendaient que les populations allaient s'habituer à la gratuité et qu'il serait impossible dans le futur de les vendre. D'autres faisaient valoir que certains refuseraient d'acheter un produit qui a été distribué gratuitement à d'autres groupes de la population. Pascaline Dupas a donc mis au point une nouvelle expérience pour tester ces hypothèses. 644 foyers sélectionnés se sont vus remettre un bon de réduction sur l'achat d'une moustiquaire, alors que celles-ci sont gratuites dans les maternités. Le montant des réductions variait de 0 à 250 shillings. Il s'avère alors que 98 % des personnes qui bénéficient d'une moustiquaire gratuite vont la chercher, contre 50 % pour 50 shillings et 11 % pour un coût compris entre 190 et 250 shillings. L'élasticité-prix se confirme par conséquent. Il apparaît en outre que l'utilisation demeure aussi forte, que l'objet soit gratuit ou non.

Un an plus tard, les chercheurs revenaient sur place pour distribuer des bons de réduction moins intéressants pour certains, puisque le prix des moustiquaires est fixé pour tous à 100 shillings. Il s'agissait ainsi de vérifier si les personnes qui avaient obtenu la première moustiquaire gratuitement étaient moins susceptibles que les autres d'en acheter une à 100 shillings. Or il s'avère que l'expérience de la gratuité ne décourage en rien. Un effet de contagion apparaît également. Ainsi, si tout l'entourage d'une personne a obtenu une moustiquaire gratuite, celle-ci est prête à payer pour l'acheter. Il est possible d'y voir un effet d'apprentissage, dans la mesure où il est possible d'en mesurer les bénéfices.

En conclusion, les évaluations expérimentales sont devenues un outil de recherche important en économie du développement. Elles permettent non seulement d'évaluer après coup les actions réalisées, mais, plus encore, de développer des approches nouvelles. Elles

influencent directement les politiques. Dans le domaine du déparasitage, il est apparu au travers d'études que cette solution permettait de lutter d'une façon très efficace pour un coût très faible contre l'absentéisme scolaire. Le Kenya a donc décidé d'implémenter une politique de déparasitage au niveau national, tandis que le Yémen et l'Éthiopie y travaillent. Une entreprise pharmaceutique a décidé de donner 300 millions de pilules à des gouvernements et à des entreprises privées s'engageant dans les actions de déparasitage. Par ailleurs, au Kenya, les études menées sur l'utilisation des moustiquaires ont abouti à une distribution gratuite dans les maternités. Les résultats de ces études sont publiés largement, ce qui éclipse le débat sur les méthodes et recentre celui-ci sur des questions de fond. Dans le domaine des affaires sociales ou de la santé, se dessinent de même de nombreuses pistes prospectives.

SYNTHÈSE DES DÉBATS AVEC LA SALLE

Problèmes éthiques liés au tirage au sort

Question de la salle

Je suis étonnée de constater que rien n'a été dit sur le plan moral sur le principe du tirage au sort des populations qui bénéficient d'une politique. En France, nous pouvons penser à l'accompagnement à la recherche d'un emploi, par le secteur privé, le secteur public traditionnel ou le secteur public renforcé. Ce principe pose-t-il des problèmes éthiques aux chercheurs ?

Réponse de Mme Duflo

Je pense qu'il faut séparer l'aspect éthique de l'aspect pratique.

Il faut d'abord préciser que toutes les expériences menées par des chercheurs dépendant d'universités américaines ont toutes été encadrées par des processus éthiques identiques à ceux qui gouvernent les protocoles cliniques. Ils permettent de s'assurer de l'absence de traitements dégradants ou de risques pour les personnes choisies. Ainsi, un comité d'éthique détermine si les règles ont été respectées. Pour la plupart des projets, au Kenya par exemple, un autre comité d'éthique est mis en place dans le pays où est menée l'expérimentation. Cependant, il est difficile de répondre à la question : l'aléatoire est-il réellement éthique ? Parfois, c'est le cas. Dans d'autres cas, non. Par exemple, une ONG proposait une expérimentation sur un programme de renutrition d'enfants. J'ai considéré, en tant que chercheur, qu'une telle étude était contraire aux principes éthiques, dans la mesure où les enfants du groupe test voyaient leur survie mise en jeu. Cependant, dans la plupart des cas, de tels problèmes ne se posent pas, dans la mesure où les budgets des partenaires avec lesquels nous travaillons s'avèrent extrêmement limités. Leurs programmes excluent de fait de larges populations. Il serait au contraire bien peu éthique de ne pas profiter de ces situations de pénurie pour entreprendre une étude qui pourrait amener à la généralisation du programme. Bien souvent, plusieurs programmes différents sont mis au point. Ils sont tous évalués comparativement, avant que l'un d'eux ne soit choisi pour généralisation.

Sur le plan pratique, nous croyions que nous allions nous heurter à des réticences sur le terrain. Il semblerait que cela ne soit pas le cas. Les personnes qui participent à l'expérimentation sont informées sur le but poursuivi. De manière générale, elles s'avèrent très sensibles aux arguments développés par les chercheurs. Peut-être l'habitude de l'arbitraire, aux États-Unis comme au Kenya, joue-t-elle également. Le hasard est alors considéré comme une méthode d'allocation assez juste, dans un contexte de ressources rares, de népotisme ou de corruption. En France, en ce qui concerne l'expérimentation menée sur l'accompagnement des chômeurs, il semblerait qu'après de fortes réticences initiales, l'expérience ait été ensuite assez bien acceptée. Il faut cependant tenir compte du fait que le nombre de places dans le système ANPE renforcé était limité par les budgets. Il n'était alors pas question de développer un programme universel. Il est en revanche très complexe de lancer un programme « randomisé » lorsque le public s'attend à ce qu'il soit universel. C'est un effort pédagogique qu'il faut alors mener. En tout cas, dans mon travail, je n'ai jamais été véritablement confrontée à une opposition fondamentale, aux États-Unis comme au Kenya.

Durée optimale pour une évaluation

Question de la salle

À ce propos, existe-t-il également une durée optimale, dans la mesure où certaines politiques sont censées produire des effets sur le long terme ? Parfois, il est impossible d'expérimenter à l'avance. C'est le cas notamment pour la question de la formation des médecins, qui est très longue.

Réponse de Mme Duflo

En ce qui concerne l'évaluation aléatoire prospective, c'est la durée du programme qui détermine la durée de l'évaluation. Ainsi, pour reprendre l'exemple de la formation des médecins, l'impact n'est connu qu'après un temps assez long. Faut-il en conclure que l'évaluation ne sert à rien ? Non, car dès le moment où le résultat est connu, il est possible de revoir la politique mise en place. Imaginons une réforme de l'internat. Il est possible d'imaginer une mise en place aléatoire dans certains CHU dans un premier temps une première année, avant de l'étendre au reste du territoire. Ainsi, se dessine une cohorte intéressante. Il s'agit de celle des jeunes internes qui ont, la même année, été formés ou non selon les principes de la nouvelle réforme. Cette cohorte peut être suivie d'année en année. Il est probable que le suivi de ces cohortes donnera des résultats intéressants dans un premier temps, même s'il ne répond pas à l'intégralité des questions. Dans la mesure où les politiques suivent souvent un cycle, il n'est pas impossible que les réponses qui seront obtenues sur le long terme s'avèrent pleinement pertinentes dans le futur.

Cependant, il n'est pas toujours possible d'établir de telles cohortes. Par exemple, il est possible d'étudier pendant deux ans une communauté avant qu'elle n'accède à la possibilité de contracter un microcrédit. Il est en revanche impossible de distinguer une population qui y a accès d'une population qui n'y a pas accès. Lorsqu'il existe une cohorte, il est possible de faire avancer l'évaluation en parallèle de la mise en œuvre de la politique.

Paramètre optimal d'une évaluation

Question de la salle

En tant que membre d'une ONG, j'ai eu l'occasion de participer à une instance d'évaluation qui n'a pas abouti, dans le domaine de l'insertion par l'activité économique. Le volume et la durée des études étaient véritablement trop importants, alors même qu'il ne s'agissait pas d'un sujet véritablement conflictuel. Je me demande donc s'il existe un périmètre optimal pour une évaluation.

Réponse de Mme Fouquet

Je suis étonnée de votre remarque sur l'évaluation de l'insertion par l'activité économique. Il me semblait que ce sujet avait donné lieu à un rapport dans le cadre de l'activité du Comité national de l'évaluation, qui était en place entre 1998 et 2001. Certains ont pu croire que le non-renouvellement du mandat du comité avait annulé les travaux en cours, alors que ce n'était pas le cas. Ils ont donné lieu à des rapports publiés à la Documentation française ; mais sans aucune publicité ni soutien à la diffusion, ces rapports sont restés confidentiels.

Beaucoup ont conservé un mauvais souvenir de cette période, en raison de la durée des évaluations mises en place. S'agissant d'un comité interministériel, tous les sujets d'évaluation choisis devaient dépendre de plusieurs ministères et s'avéraient au final particulièrement complexes.

Dans le cas particulier de la politique de sécurité routière et du respect des limitations de vitesse, l'évaluation a duré quatre ans. Tout d'abord il a fallu déterminer les ministères concernés : Intérieur (police), Défense (gendarmerie), Finances (recouvrement des amendes), Justice, Éducation. Chaque ministère avait une vision très différente du sujet. Il aura fallu 18 mois pour s'accorder sur la composition de l'instance. Ensuite, six mois auront été nécessaires pour déterminer quelles questions évaluer. Un appel d'offres a ensuite été lancé. Afin d'éviter que l'organisme chargé de la réalisation de l'évaluation ne soit considéré comme juge et partie, c'est un cabinet d'audit qui ne connaissait rien à la question ni au métier d'évaluateur qui a été retenu, au détriment de chercheurs et d'universités qui avaient l'habitude de travailler avec l'un ou l'autre de ces ministères. Il allait de soi que les autres ministères disqualifiaient un par un ces prétendants.

Le cabinet d'audit a suivi pas à pas le circuit des contraventions pour comprendre pourquoi les deux tiers d'entre elles restaient impayées. Le rapport est très intéressant par son analyse du système d'acteurs : personne n'était responsable de cet état de fait, mais tous l'étaient collectivement. Avant de passer aux Finances, certaines contraventions étaient supprimées par la Police qui s'en servait de monnaie d'échange pour ses informateurs (« les cousins »). Aux Finances, les circuits de recouvrement avec leurs différentes étapes de pénalisation pour impayés duraient plus de 9 mois. La Justice, surchargée, considérait avoir des affaires plus importantes à traiter que les infractions au code de la route. Pour finir, l'amnistie du 14 juillet venait régulièrement balayer les affaires qui s'empilaient. L'étude avait été réalisée au temps où Jean-Claude Gaysot était ministre des Transports, mais le rapport a été remis en 2002 à Nicolas Sarkozy, qui a décidé de suivre les conclusions de l'étude, qui préconisait d'automatiser le système. Il est remarquable qu'une telle étude ait survécu à l'alternance politique, alors même que se posaient d'importants problèmes à l'origine sur la façon d'envisager la question.

En conclusion, c'est l'aspect interministériel qui a pénalisé le fonctionnement de cette institution. Le fait de choisir une question ne relevant que d'un seul ministère simplifie considérablement la question. Il est possible de restreindre le périmètre pour n'étudier qu'un seul dispositif. Mais c'est alors prendre le risque de ne pas repérer les incohérences avec d'autres dispositifs proches.

Sur les sujets très conflictuels, il est difficile de réaliser une bonne évaluation. Dans ce cas, il faut d'abord comprendre sur quoi portent les désaccords. Ce n'est qu'ensuite qu'il est possible de dépasser ce stade. Les décideurs politiques sont toujours très pressés. C'est la raison pour laquelle il est préférable de laisser la question de la durée de l'évaluation à la main de l'administration centrale. Au niveau territorial, les élus sont proches de leur administration, pour ceux qui sont directement en charge de l'exécutif. Ceux-ci sont par conséquent en capacité de mettre en place des suivis évaluatifs avec des parties relevant de l'acquisition de connaissances (comme le suivi de cohortes par échantillon), à mesure que se met en place un dispositif. Un temps peut être réservé pour une analyse plus large du processus. Le Conseil régional Rhône-Alpes vient de mettre en place un comité scientifique de suivi des évaluations. Ce dernier a décidé de fixer à 18 mois la durée de base des évaluations, tout en faisant des « suivis évaluatifs » à 4 ou 6 mois. C'est un bon compromis. L'existence de données nationales sur les champs qui sont explorés peut faire gagner beaucoup de temps à la démarche d'évaluation.

Pays en avance au niveau de l'évaluation

Question de la salle

Certains pays dans le monde ont-ils pris de l'avance dans le domaine de l'évaluation des politiques publiques ? La France pourrait-elle s'inspirer de l'exemple de certains ?

Réponse de Mme Fouquet

Il faut distinguer les pays nordiques des pays latins en matière d'évaluation de politique publique. Lors du colloque sur l'évaluation des politiques publiques en Europe⁴, il est apparu une fracture. Dans les pays latins, il semblerait que l'évaluation soit associée à la notion de culpabilité. Ceci constitue un frein notable au développement de cette notion. Il s'agit pourtant d'une erreur de fond. L'évaluation n'est pas destinée à désigner un coupable, mais à déterminer comment l'action collective peut être menée de façon plus efficace ou plus efficiente.

Les pays nordiques et anglo-saxons, qui ont l'habitude d'un débat public apaisé, ont en revanche adopté une approche plus saine. En Grande-Bretagne, par exemple, l'évaluation législative est très largement portée par le *National audit office*, qui certifie les comptes et réalise des évaluations sous forme d'enquêtes approfondies. La Suède reste très friande d'évaluation. L'administration suédoise se compose de ministères employant une quarantaine de personnes et gérant chacun diverses agences. L'administration centrale donne des objectifs aux agences et un contrôle de gestion et une évaluation s'ensuivent. L'évaluation entre dans les missions premières de l'État.

Aux États-Unis, l'évaluation est portée par deux organismes. L'un est propre à l'exécutif et l'autre propre au législatif. Chacun surveille l'autre et le travail fourni par ces deux organismes demeure de très bonne qualité.

Effet du développement de l'évaluation sur les programmes de statistique publique

Question de la salle

Quel effet aura le développement de l'évaluation sur les programmes de statistique publique ? Comment rassembler des expériences ponctuelles pour créer un véritable appareil statistique et systématique ? Je m'interrogeais par ailleurs sur l'opportunité d'externaliser le processus d'évaluation.

Réponse de Mme Fouquet

L'existence d'un Comité ministériel chargé d'établir un programme pourrait ainsi donner des indications aux directions d'études (DARES, DREES) ou à l'IGAS pour anticiper sur les thèmes qui vont être abordés. Plus généralement, l'INSEE s'est progressivement éloigné des préoccupations des décideurs politiques, laissant ce soin aux services statistiques ministériels. Je ne pense donc pas que cette nouvelle organisation aura un impact direct sur l'INSEE. En revanche, les programmes de travail des directions d'études des ministères peu-

4. Organisé dans le cadre de la Présidence française de l'Union européenne en juillet 2008, au Parlement européen de Strasbourg, par la SFE et la DeGEval (société d'évaluation de langue allemande) cf. Fouquet, Measson, 2009.

vent évoluer, d'autant que les enquêtes de fond menées par ces directions d'études ont pour objectif d'aider à la décision.

Toute évaluation a pour but l'analyse d'impact. La méthode expérimentale me paraît particulièrement appropriée pour des projets précis. En revanche, sur les programmes complexes, il est difficile de mettre en place un système d'expérimentation efficace à l'avance. Par ailleurs, il existe des méthodes d'analyse qui permettent d'imaginer ce qui se serait passé si la politique n'avait pas été implémentée. Il existe une multitude de méthodes économétriques, d'appariement des échantillons *ex post* ou qualitatives. À ce propos, je dois évoquer un séminaire qui a eu lieu le 30 avril 2009 sur la pluralité des méthodes disponibles pour les évaluations d'impact, parmi lesquelles les méthodes expérimentales.

En ce qui concerne le débat sur l'externalisation, c'est la question de l'indépendance de l'évaluateur qui se pose. C'est cette notion qui est désignée par la charte de la SFE comme la « distanciation » par rapport à ceux qui sont chargés de mettre en place une politique ou qui la décident. La conduite d'une évaluation nécessite en tout premier lieu de déterminer qui est concerné par la politique. Il faut alors prendre en compte le point de vue de tous les acteurs, décideurs, financeurs, opérateurs ou bénéficiaires. Les instances d'évaluation tiennent à distance dans la mesure du possible les décideurs et les opérateurs, même si ces derniers se doivent de suivre le cheminement de la réflexion, qui aboutit à des recommandations.

Souvent, les évaluateurs se voient dire au moment de la remise de leur rapport que celui-ci est inutile, dans la mesure où la décision a déjà été prise. Il s'avère généralement que les décisions vont justement dans le sens des recommandations. Ce paradoxe apparent s'explique par le fait que les travaux ne sont pas secrets et les décideurs sont tenus au courant, de manière régulière de l'avancement des réflexions. Il faut savoir préserver l'osmose, tout en maintenant la distanciation.

BIBLIOGRAPHIE

- Bannerjee A., Duflo E., Glennester R., Kothari D.**, 2010, «Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives», *British Medical Journal*, 340: c2220.
- Bannerjee A., Duflo E.**, 2008, *The experimental approach to development economics*, Massachusetts Institute of Technology, Department of Economics and Abdul Latif Jameel Poverty Action Lab
- Beaman L., Chattopadhyay R., Duflo E., Pande R., Topalova T.**, 2008, *Powerful Women: Does Exposure Reduce Bias?*, NBER Working Papers 14198, National Bureau of Economic Research, Inc.
- Braconnier R., Cauquil G.**, 2010, *L'évaluation des politiques publiques - Pour une recherche d'efficacité*. Préface de Michel Rocard, Éditions CNDP - ESEN, 2010.
- Cauquil G., Lafore R.** (dir), 2006, *Évaluer les politiques sociales*, SFE/L'Harmattan, 335 p.
- Chattopadhyay R., Duflo E.**, 2004, «Women as policy makers: Evidence from a randomized policy experiment in India», *Econometrica*, vol. 72, n° 5, 1404-1443
- Cohen J., Dupas P.**, 2009, «Free Distribution or Cost-Sharing Evidence from a Randomized Malaria Prevention Experiment», *Quarterly Journal of Economics*.
- Conseil scientifique de l'évaluation**, 1996, *Petit guide de l'évaluation de politiques publiques*.
- Duflo E., Kremer M.**, 2008, *Use of randomization in the evaluation of development effectiveness*, World Bank Operations Evaluation Department
- Duflo E.**, 2006, *Field experiments in development economics*, Paper for the World Congress of the Econometric Society
- Duflo E.**, 2004, *Scaling up and evaluation*, Annual World Bank Conference on Development Economics
- Dupas P.**, *Short-Run Subsidies and Long-Term Adoption of New Health Products: Experimental Evidence from Kenya*, Forthcoming
- Fouquet A., Measson L.** (dir), 2009, *L'évaluation des politiques publiques en Europe. Cultures et futurs. Policy and Programme Evaluation in Europe. Cultures and Prospects*, ouvrage bilingue, SFE/L'Harmattan, 440 p.
- Fouquet A.**, 2007, «L'évaluation des politiques publiques, indispensable outil de gouvernance dans une société complexe», in **Y. Cannac et S. Trosa** (dir), *La réforme dont l'État a besoin. Pour un management par la confiance et la responsabilité*, Dunod.
- Fouquet A.**, 2010, «L'usage des statistiques: de l'aide à la décision à l'évaluation des politiques publiques», *Revue Française des Affaires Sociales*, janvier.
- Jacot H., Fouquet A.**, 2007, *Le citoyen, l'élu, l'expert. Pour une démarche pluraliste d'évaluation des politiques publiques*, SFE/L'Harmattan, 241 p.
- Kremer M., Miguel E., Thornton R.**, 2009, «Incentives to learn», *Review of Economics and Statistics* (Forthcoming).
- Mouterde F., Trosa S.**, 2010, *Les nouvelles frontières de l'évaluation. 1989-2009: vingt ans d'évaluation en France, et demain?* texte d'ouverture de Philippe Seguin, SFE/L'Harmattan. 340 p.

- Perret B.**, 2001, *L'évaluation des politiques publiques*, coll. « Repères », La Découverte. 125 p.
- Revue française des affaires sociales**, *Évaluer les politiques sociales*, numéro coordonné par P. Duran, vol. 1-2, 2010.
- Société française de l'évaluation**, *Charte de l'évaluation des politiques publiques et des programmes publics*; <http://www.sfe-asso.fr>
- Société française de l'évaluation et Société française de statistiques**, 2010, *Statistiques et évaluation, quelle valeur (appréciation) donner à une valeur (mesure)*, Les cahiers de la SFE, www.sfe-asso.fr.
- Société française de l'évaluation**, 2009, « La notion d'utilité sociale au défi de son identité dans l'évaluation des politiques publiques », www.sfe-asso.fr
- Ternier M., Phelippeau E., Malibert P., Vilmart C.**, *La politique de sécurité routière: les systèmes locaux de contrôle-sanction*, Conseil national de l'Évaluation, La Documentation française, Paris.
- Thornton R.**, 2008, « The Demand for, and Impact of, Learning HIV Status », *American Economic Review*, 98 (5), 1829-63.

L'IDENTIFICATION DE L'EFFET PROPRE D'UN PROGRAMME

La deuxième séance du séminaire sur les méthodes d'évaluation de politiques publiques était consacrée aux méthodes économétriques d'évaluation *ex post* de l'impact d'un programme sur ces bénéficiaires. Elle était articulée autour de deux interventions :

- Bruno Crépon, directeur de recherches au Centre de recherches en économie et statistique (CREST), a présenté le cadre général d'analyse, les méthodes d'appariement et les expériences contrôlées ;
- Pauline Givord, chef du Bureau « Marché du travail et stratégies d'entreprises » à l'Institut national de la statistique et des études économiques (INSEE), a présenté les méthodes de différences de différences, de variables instrumentales et de régressions sur discontinuités.

Le cadre général d'analyse, les méthodes d'appariement et les expériences contrôlées

Bruno Crépon (CREST-INSEE)

L'évaluation d'impact

Évaluation de processus et d'impact

Par évaluation, on comprend souvent évaluer le processus de mise en œuvre d'un programme généralement destiné à une population spécifique. Les analyses s'apparentent alors à un audit : compter le nombre de bénéficiaires effectifs, déterminer le taux d'entrée dans le programme, vérifier que les services prévus ont bien été offerts, qu'ils étaient de bonne qualité, etc. Ce sont des questions importantes et utiles, mais pas suffisantes.

Les programmes ont en effet surtout pour finalité d'atteindre un objectif, par exemple favoriser la réinsertion professionnelle des personnes privées d'emploi et les évaluations de processus ne donnent aucune information sur ce point qui est soit ignoré, soit considéré comme acquis. Il est pourtant crucial de vérifier si les objectifs assignés aux programmes sont atteints et si ces programmes apportent une réelle plus-value. En outre, il est également utile de déterminer si le programme est rentable ou non à partir d'une analyse coût-bénéfice et d'explorer des pistes d'amélioration. C'est l'objet des évaluations dites « d'impact ».

Le biais de sélection : un problème central

La formalisation de ce que peut être l'impact d'un dispositif, au niveau individuel ou collectif, et la façon dont on peut le mesurer ont été l'objet de recherches importantes (Angrist et Krueger, 2001 ; Heckman et Vytlačil, 2005). L'effet causal de la participation à un dispositif (par exemple la sortie du chômage vers l'emploi) est défini comme la comparaison entre la situation d'un individu bénéficiaire et ce qu'aurait été sa situation s'il n'avait pas participé au programme. On compare donc une situation réelle et observée, celle de l'individu bénéficiaire, avec une situation virtuelle et inobservée. L'effet causal a une caractéristique centrale : il est inobservé. On n'observe pour chaque individu qu'une seule des deux situations : avec et sans participation au programme. Il y a nécessairement une des deux informations qui est manquante. Pour mesurer l'impact d'un programme, il est nécessaire de reconstruire cette situation inobservée. L'idée centrale est de s'appuyer sur la situation observée pour des indi-

vidus n'ayant pas participé au programme. On est alors amené à confronter la situation moyenne observée pour un groupe de participants et la situation moyenne observée pour un groupe de non-participants – un groupe de contrôle. La question est alors celle du choix de ce groupe de contrôle.

Prenons l'exemple de l'évaluation des programmes « Opérateurs privés de placement » de l'ASEDIC et « Cap vers l'entreprise » de l'ANPE, réalisée entre 2007 et 2009. Ces programmes ciblent spécifiquement les demandeurs d'emploi pour qui on a identifié un risque de chômage de longue durée lors de leur inscription à l'ANPE. L'idée est de les faire bénéficier d'une prise en charge plus intense en les adressant à des conseillers qui ont moins de cas à traiter simultanément : alors que le portefeuille « normal » d'un agent de l'ANPE est de 120 demandeurs d'emploi, il est de 60 demandeurs d'emploi pour les conseillers participant au programme. L'évaluation d'impact peut alors s'intéresser à l'effet de la réduction de la taille du portefeuille par conseiller sur le retour à l'emploi. Une méthode naturelle consiste à choisir les non-bénéficiaires comme groupe de contrôle. Cette méthode est simple mais certainement erronée car comme on l'a vu la population éligible est une population spécifique dont le retour à l'emploi est certainement plus difficile en l'absence de programme. De la même façon, choisir comme groupe de contrôle les individus éligibles qui ont choisi de ne pas participer au programme peut conduire à des résultats biaisés : on peut penser par exemple que ces individus refusent de participer au programme car ils ont des raisons de penser qu'ils peuvent retrouver un emploi relativement rapidement même sans participer au programme (par l'intermédiaire de leur réseau par exemple).

Ainsi, les deux populations de participants et de non-participants ne sont pas identiques, en l'absence même du programme. De ce fait, la comparaison de la situation moyenne des deux populations de bénéficiaires et de non-bénéficiaires mêle deux effets indiscernables. Le premier est l'impact du programme. C'est ce que l'on cherche à mesurer. Le deuxième est « l'effet population » qui résulte des différences préexistantes entre les deux populations. Cet effet vient contaminer la mesure à laquelle on s'intéresse et biaise l'appréciation que l'on a de l'impact de la participation au programme. Il s'agit là d'un problème fondamental de l'évaluation : « le biais de sélection ».

Il existe un large éventail de méthodes pour corriger les comparaisons que l'on effectue entre bénéficiaires et non-bénéficiaires de ce biais de sélection. Celles-ci reposent sur des hypothèses spécifiques qui ont un impact considérable sur les résultats. Ces méthodes sont en général peu robustes et font appel à de l'économétrie compliquée. Les résultats ont ainsi du mal à peser sur le débat public en raison d'un déficit de crédibilité. Le débat porte alors plus sur les méthodes et hypothèses sous-jacentes retenues pour réaliser les estimations que sur leurs résultats. Deux méthodes, la méthode dite de l'appariement et la méthode dite de l'expérimentation aléatoire permettent de lever cette difficulté.

L'appariement et l'expérimentation aléatoire

L'une des méthodes couramment employée pour l'évaluation est la méthode dite de « l'appariement » (Imbens, 2000 ; Caliendo et Kopeinig, 2005). Pour chaque individu bénéficiant du programme, il s'agit de trouver un individu n'ayant pas bénéficié du programme et ayant des caractéristiques observables similaires – sexe, âge, diplôme, historique sur le marché du travail, etc. – et de comparer leurs situations respectives. Cette méthode réduit les erreurs que l'on peut commettre en faisant une comparaison grossière, mais un biais subsiste. On ne corrige avec cette méthode que la sélection dite « observable ». Il reste la sélection dite « inobservable ». Par exemple, dans une population de chômeurs, il y a des demandeurs d'emploi prêts à faire tous les efforts possibles pour retrouver rapidement un emploi. D'autres, au contraire, sont moins motivés et peuvent éprouver une forme de découragement. Aucun fichier statis-

tique ne contient ces informations « inobservables ». Elles sont pourtant très importantes car elles conditionnent à la fois l'entrée dans un programme et la capacité à sortir spontanément du chômage vers l'emploi. Elles vont donc fortement contribuer au résultat en venant biaiser les comparaisons entre participants et non-participants. Il ne s'agit pas là d'un problème de second ordre que l'on pourrait négliger en première approximation. Heckman, Ichimura et Todd (1998) montrent que les biais résiduels avec la méthode de l'appariement sont en général de l'ordre de grandeur de l'effet du programme lui-même. En outre, les résultats obtenus sont sensibles au choix des variables sur lesquels on apparie les individus, et on se trouve alors face à une multiplicité de résultats. Ainsi, même si la méthode de l'appariement est une très bonne méthode, là où l'on a besoin d'évidence forte, son apport est limité.

Les évaluations dites « par expérimentation » ne présentent pas cet inconvénient. Dans leur version standard, ces évaluations consistent à partager de façon aléatoire la population éligible en deux. Une seule des deux populations, le groupe test, bénéficie du programme. L'intérêt de cette démarche est qu'elle permet d'obtenir deux populations statistiquement identiques : même proportion de moins de trente ans, de femmes, de diplômés, mais aussi même proportion de personnes motivées pour améliorer leur situation ou, au contraire, découragées. Au terme d'un échantillonnage aléatoire, n'importe quelle caractéristique est identiquement représentée entre les deux groupes : on est en présence d'un groupe de contrôle idéal. Une fois le programme mis en œuvre, la seule différence entre le groupe test et le groupe de contrôle ne peut provenir que du programme lui-même. Il n'y a pas d'effet population et l'effet du programme peut être mesuré en toute transparence par la différence de situations entre les deux populations. L'idée reste donc la même que dans les études par appariement, à la différence notable que l'on construit au préalable deux populations statistiquement identiques et qu'elles sont exposées différemment au programme. On n'a donc pas à reconstruire *ex post* ces deux populations, qui ne seront statistiquement identiques que sur un nombre choisi et donné de caractéristiques. Cette méthode est donc à la fois plus fiable et plus transparente. Elle est également précise lorsque les échantillons constitués sont suffisamment importants.

Mise en œuvre pratique de l'évaluation par expérimentation aléatoire

Cette idée d'évaluation par expérimentation peut paraître complexe et peu opérationnelle. Pourtant, il faut plutôt la considérer comme un principe général qu'il est possible d'insérer, le plus souvent de façon assez souple, dans le processus de mise en œuvre d'un programme. Nous présentons trois exemples types de schémas possibles de mise en œuvre d'évaluations par expérimentation.

Un tirage aléatoire direct (loterie) peut parfois paraître injuste. Pourtant, dans de nombreuses situations, le nombre de places disponibles dans le programme est limité et largement inférieur à la population éligible, en particulier pour des programmes mis en place à une petite échelle. Un tirage aléatoire peut alors sembler la seule méthode offrant une probabilité d'entrer dans le programme égale à tous les candidats. Les programmes « Opérateurs privés de placement » (OPP) et « Cap vers l'entreprise » (CVE) étaient par exemple destinés à 40 000 demandeurs d'emploi entre janvier 2007 et mars 2008 mais les bénéficiaires potentiels étaient plus de 250 000 au cours de cette période. La mise en place d'un outil informatique (Outil de constitution des cohortes) a permis d'orienter chaque demandeur d'emploi avec des risques de chômage de longue durée de façon aléatoire vers l'accompagnement OPP, l'accompagnement CVE ou le programme classique d'accompagnement de l'ANPE tout en constituant trois populations statistiquement identiques.

Certains programmes sont également mis en œuvre progressivement car on ne pourrait mobiliser les moyens nécessaires à leur développement immédiat dans toutes les zones concernées. Dans ce cas, il est possible de découper l'ensemble du territoire concerné par le

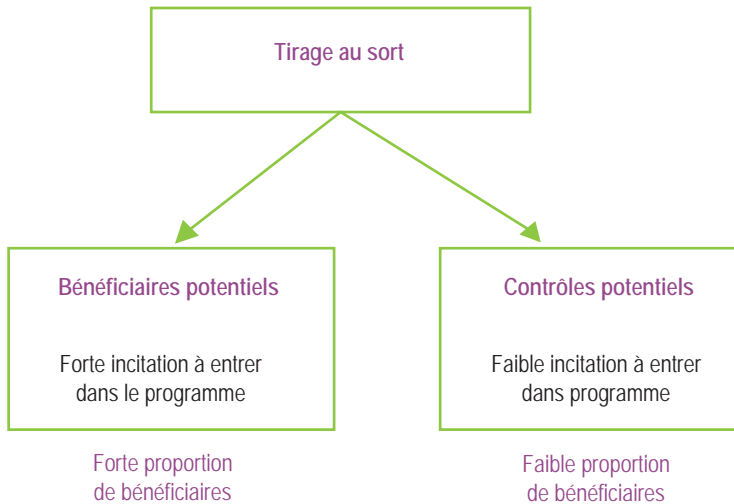
programme en un grand nombre de périmètres homogènes. Au lieu de favoriser les citoyens dont les élus sont les plus influents, on peut alors tirer au sort l'ordre selon lequel les périmètres entreront dans le programme. Imaginons par exemple un programme consistant à doter les Missions locales d'une consultation médicale afin d'améliorer l'état de santé des jeunes dans les banlieues défavorisées. Ce programme ne peut pas être développé en une année mais sur deux ans. On peut alors tirer les Missions locales qui vont entrer dans le programme dès la première année : cela constitue le groupe de traitement. Le groupe de contrôle est l'ensemble des Missions locales qui vont développer le programme la deuxième année. Il est important de remarquer qu'on ne tire alors pas des individus au sort mais des lieux ; cela est possible s'il y a un nombre suffisamment important de Missions locales.

L'«encouragement design»

Enfin, il convient de remarquer qu'en pratique, on est rarement dans la situation caricaturale de l'affectation autoritaire au groupe de traitement et au groupe de contrôle. Les bénéficiaires potentiels peuvent en effet toujours refuser d'entrer dans le programme, alors qu'à l'opposé les personnes qui ne sont pas orientées vers celui-ci (contrôles potentiels) peuvent en définitive relever du programme. Dans ce cas, on ne tire au sort que les personnes orientées ou non vers le programme et on leur envoie un signal respectivement élevé ou faible pour qu'elles acceptent ou refusent d'entrer dans le programme (schéma 2).

SCHÉMA 2

L'«encouragement design»



Prenons l'exemple de l'évaluation des programmes OPP et CVE. Parmi les 50 563 demandeurs d'emploi orientés vers le programme OPP, seuls 47% ont accepté d'entrer dans le programme, 33% ont refusé explicitement, 20% n'ont pas répondu et 0,3% ayant eu connaissance du programme CVE ont demandé à participer à ce dernier programme. Concernant le programme CVE, sur les 9 938 personnes auxquelles le programme a été proposé, 56,3% ont refusé ou n'ont pas donné de réponse et 0,9% a demandé de bénéficier du programme OPP (tableau 2)

TABLEAU 2

Taux d'acceptation des programmes opérateurs privés de placement (OPP) et cap vers l'entreprise (CVE) pour les personnes du flux indemnisable

OPP	Effectif orienté	Part de signataires (en %)	Part de non-signataires (en %)		
			Refus explicite	Pas d'information	En CVE
	50 563	47	33	20	0,3
CVE	Effectif orienté	Part des personnes ayant accepté (en %)	Part des personnes n'ayant pas accepté (en %)		
			Refus ou raison inconnue		En OPP
	9 938	42,8	56,3		0,9

Lecture • Sur les 9 938 demandeurs d'emploi auxquels le programme CVE a été proposé, 42,8% ont accepté d'entrer dans ce programme.

Champ • Demandeurs d'emploi qui se sont inscrits au chômage en 2007 dans l'une des 393 agences locales de l'emploi participant à l'expérimentation et ont été identifiés comme présentant un risque de chômage de longue durée.

Le calcul de l'effet de la plus-value des deux programmes par rapport à l'accompagnement classique est toujours possible. Néanmoins, seules les trois sous-populations orientées respectivement vers le programme OPP, le programme CVE et l'accompagnement classique sont comparables. En revanche des différences existent entre les sous-populations relevant en définitive des différents programmes. Ainsi, les personnes effectivement entrées dans le programme OPP sont en moyenne plus âgées et plus éduquées que les personnes qui y ont été orientées (tableau 3). Le refus d'entrer dans un programme est en effet lié en général à des caractéristiques personnelles.

TABLEAU 3

Statistiques descriptives concernant la population participant à l'expérimentation

Flux indemnisable, zones OPP	Orienté		Entrant
	Classique (1)	OPP (2)	OPP (3)
Diplôme supérieur	0,24	0,25	0,25
Bac	0,18	0,18	0,19*
CAP/BEP, autres secondaires	0,39	0,39	0,40*
BEPC ou non-diplômés	0,18	0,19	0,17*
Âge	36,86	36,77	37,16*
Femme	0,51	0,50	0,51
Ouvrier	0,16	0,15*	0,15
Employé	0,61	0,62	0,61*
Tecnicien	0,10	0,10	0,11*
Cadre	0,12	0,12	0,12

Lecture • L'âge moyen des demandeurs d'emploi dans la population orientée vers le programme OPP était de 36,77 ans et n'était pas statistiquement différent de l'âge moyen des demandeurs d'emploi dans la population orientée vers l'accompagnement classique à la différence de l'âge moyen des demandeurs d'emploi entrés dans le programme OPP (37,16 ans).

Champ • Demandeurs d'emploi inscrits au chômage en 2007 dans l'une des 393 agences locales de l'emploi participant à l'expérimentation et identifiés comme présentant un risque de chômage de longue durée.

Pour évaluer l'effet du programme, on fait l'hypothèse que la différence observée entre le taux de retour à l'emploi (\bar{Y}) des populations orientées vers le programme CVE et celui des populations orientées vers l'accompagnement classique ne peut être due qu'à la participation au programme d'une partie des individus qui y ont été orientés. On divise donc cette différence par la proportion d'acceptants :

$$Effet_{CVE} = \frac{\bar{Y}^{orienté CVE} - \bar{Y}^{orienté CVE}}{Pr op_{Acceptation CVE}}$$

L'effet du programme CVE ne peut en effet être calculé que sur des populations aux caractéristiques initiales comparables, à savoir les personnes orientées vers le programme CVE⁵ ou vers l'accompagnement classique et non entre les personnes entrées dans le programme CVE ou dans l'accompagnement classique et participant au programme CVE. Cette approche permet en fait de donner de la robustesse aux évaluations par expérimentation. Il leur est en effet fréquemment reproché d'être des cadres trop contraignants et de ne proposer des résultats que lorsque le protocole a été scrupuleusement respecté. Dès lors que l'on s'écarterait un tant soit peu du protocole, on ne serait plus dans une situation permettant de tirer des conclusions de l'expérimentation.

En particulier, il se pourrait que, dans le cas d'une loterie, des personnes initialement affectées au groupe de contrôle veuillent absolument entrer dans le programme. On pourrait penser que ce type d'événement compromet de façon rédhibitoire la validité des résultats. Il n'en est rien, ces écarts par rapport au protocole expérimental initial sont prises en compte dans l'analyse, et on peut donc continuer à évaluer de façon fiable et rigoureuse l'efficacité du programme (Angrist et Krueger, 2001).

En conclusion, les résultats de l'expérimentation OPP-CVE sont assez marquants. Il semble en effet que l'accompagnement OPP n'a pas d'effet significatif sur le retour à l'emploi des chômeurs indemnisés, quel que soit l'horizon temporel considéré. En revanche, le programme CVE a bien un effet positif sur cette population. Cet effet se situe entre 5 et 6 points de pourcentage selon que l'on se place à 3, 6, 9 ou 12 mois (tableau 4). On trouve également un effet du même ordre de grandeur pour les chômeurs non indemnisés (tableau 5). Ces résultats sont surprenants. Ils vont en partie à l'encontre de l'idée largement répandue selon laquelle le secteur privé serait plus efficace que le secteur public, en particulier si on lui donne de bonnes incitations comme c'est le cas pour l'accompagnement OPP (l'organisme de placement reçoit une prime à chaque demandeur d'emploi placé). On peut se demander quelle crédibilité ces résultats auraient eu, quelle confiance l'opinion publique leur aurait accordé s'ils n'avaient pas été obtenus dans le cadre d'une expérimentation.

5. Que ces personnes aient ou non accepté de participer au programme.

TABLEAU 4

Effet des accompagnements sur le retour à l'emploi des chômeurs indemnisables

Accompagnements CVE et OPP dans le flux indemnisable										
	Accompagnement OPP					Accompagnement CVE				
	Est.	Std.	Inf.	Sup.	Effectif	Est.	Std.	Inf.	Sup.	Effectif
Emploi à 3 mois	0,9	(1,2)	-1,5	3,3	69 440	5,3	(1,4)	2,5	8,1	69 440
Emploi à 6 mois	3,5	(1,5)	0,5	6,5	69 440	6,3	(1,6)	3,2	9,5	69 440
Emploi à 9 mois	2,8	(1,9)	-1,1	6,7	53 235	6,3	(1,8)	2,6	10,0	53 235
Emploi à 12 mois	-0,3	(2,6)	-5,6	5,0	33 892	5,1	(2,6)	-0,1	10,3	33 892

Lecture • L'accompagnement CVE augmente de 6,3 points de pourcentage le retour à l'emploi 9 mois après l'entrée dans le programme par rapport à l'accompagnement classique. L'écart type de cette estimation portant sur 53 235 demandeurs d'emploi est de 1,8 point et la vraie valeur de l'effet de l'accompagnement CVE à 95 % de chances d'être comprise entre 2,6 et 10 points.

Champ • Demandeurs d'emploi indemnisables qui se sont inscrits au chômage en 2007 dans l'une des 393 agences locales de l'emploi participant à l'expérimentation et ont été identifiés comme présentant un risque de chômage de longue durée.

TABLEAU 5

Effet de l'accompagnement CVE sur le retour à l'emploi des chômeurs

Accompagnements CVE dans le flux non indemnisable et le stock										
	Flux non indemnisable					Stock				
	Est.	Std.	Inf.	Sup.	Effectif	Est.	Std.	Inf.	Sup.	Effectif
Emploi à 3 mois	4,9	(1,5)	1,8	8,0	55 470	2,1	(1,4)	-0,8	5,0	78 671
Emploi à 6 mois	7,0	(1,6)	3,6	10,3	55 470	3,9	(1,7)	0,5	7,4	78 671
Emploi à 9 mois	5,2	(1,9)	1,3	9,1	41 395	6,8	(1,9)	3,0	10,6	69 535
Emploi à 12 mois	6,5	(2,6)	1,3	11,8	27 065	8,9	(2,3)	4,3	13,5	56 731

Lecture • L'accompagnement CVE augmente de 5,2 points de pourcentage le retour à l'emploi 9 mois après l'entrée dans le programme par rapport à l'accompagnement classique pour les personnes non indemnisables inscrites au chômage depuis moins de 3 mois. L'écart type de cette estimation portant sur 41 395 demandeurs d'emploi est de 1,9 point et la vraie valeur de l'effet de l'accompagnement CVE à 95 % de chances d'être comprise entre 1,3 et 9,1 points.

Champ • Demandeurs d'emploi non indemnisables qui se sont inscrits au chômage en 2007 dans l'une des 393 agences locales de l'emploi participant à l'expérimentation et ont été identifiés comme présentant un risque de chômage de longue durée.

Un exemple d'expérience contrôlée : le projet CréaJeunes

Le contexte

En novembre 2007, le Haut commissariat aux solidarités actives contre la pauvreté a lancé un appel d'offres dans le cadre du Grenelle de l'insertion. Cet appel d'offres, inédit dans notre pays, était destiné à favoriser simultanément l'émergence de programmes d'insertion et leur évaluation. La principale innovation de cette démarche est d'associer, à la mise en

œuvre de nouveaux programmes, leur évaluation. Ainsi, il créait une forte incitation à s'assurer dès la conception des programmes que ceux-ci seront évaluables dans de bonnes conditions. Le centre de recherche en économie et statistiques (CREST) s'est associé au Poverty Action Lab de l'École d'économie de Paris pour répondre à cet appel d'offres et porter l'évaluation de trois programmes: un programme de prévention des ruptures dans l'apprentissage conduit par la Mission locale de Tulle, un dispositif permettant l'amélioration de l'intégration professionnelle des jeunes sortis du système éducatif sans qualification (Écoles de la deuxième chance) et ainsi un dispositif porté par l'association pour le droit à l'initiative économique (ADIE) permettant d'aider les jeunes à créer leur entreprise (CréaJeunes). Nous détaillons le procédé expérimental mis en place pour l'évaluation de ce dernier programme.

Le programme CréaJeunes est destiné aux jeunes de 18 à 30 ans issus des quartiers sensibles ayant un projet de création d'entreprises. En effet, selon un sondage réalisé en 2006, 50% des jeunes vivant en zone urbaine sensible (ZUS) se déclaraient potentiellement très intéressés par la création d'entreprise alors même que le taux de chômage des jeunes actifs dans ces quartiers est très élevé. Toutefois, s'ils manifestent un intérêt réel, les jeunes sont le plus souvent limités par leur manque d'expérience et l'absence de fonds propres.

Ainsi, CréaJeunes est un projet pilote qui a débuté en 2007. Ce programme propose un accompagnement avant et après la création, adapté au profil du porteur du projet. Cet accompagnement consiste en un coaching individuel et collectif, des modules de formation très concrets, une mise en réseau avec des entreprises du même secteur et éventuellement des immersions. La phase d'accompagnement pour la création dure de 1 à 4 mois et est suivie d'un accompagnement sur 18 mois du jeune entrepreneur. CréaJeunes offre également des possibilités de financement pour certains frais spécifiques liés à la préparation du projet (comme l'étude de marché), ainsi que pour le démarrage de l'activité. Un microcrédit peut en outre être proposé au jeune dans le cadre de sa création d'entreprise.

Le contexte pour l'expérimentation est très favorable. En effet, les dispositifs d'aide à la création d'entreprise ne sont jamais évalués de façon rigoureuse ce qui nuit à leur lisibilité et sans doute également à l'efficacité de l'action publique dans ce domaine. Le besoin d'évaluation est pourtant important, de même que le besoin d'une meilleure appréhension des mécanismes économiques à l'œuvre. Des dispositifs de ce type peuvent permettre d'insérer les jeunes et ainsi de réduire significativement les inégalités et l'exclusion. L'enjeu est donc majeur. Le démarrage d'un nouveau programme constitue une occasion unique de mettre en place un système permettant de mesurer rigoureusement son impact en intégrant les contraintes liées à l'évaluation aléatoire dès la conception du programme. Ces contraintes sont notamment la création de groupes statistiquement identiques par l'affectation aléatoire d'un nombre assez important d'individus dans les deux groupes. Il faudra ensuite être capable de suivre les trajectoires des individus des deux groupes afin de les comparer: les différences observées seront alors entièrement attribuées à la participation ou non au programme.

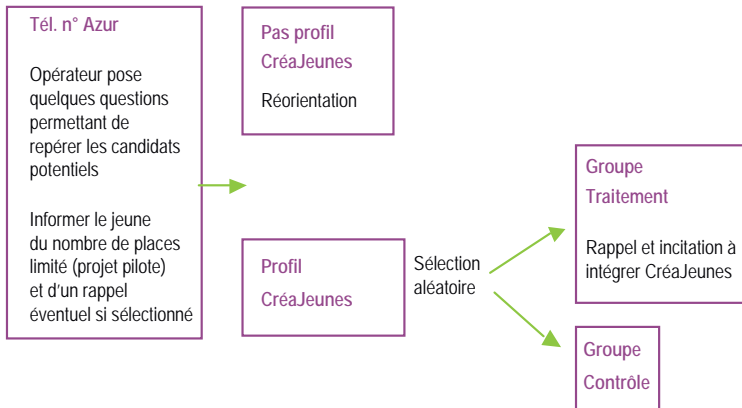
Le dispositif expérimental

Le dispositif expérimental choisi a notamment consisté à mettre en place, dès le lancement du programme, un numéro Azur (schéma 3). Lorsqu'un jeune appelle ce numéro, un opérateur lui pose des questions sur son âge, la nature de son projet et de son financement, sa disponibilité, l'interroge sur son lieu de résidence et lui demande de lui communiquer ses coordonnées. Il informe en outre la personne intéressée du fait que ce projet est un projet pilote et, de ce fait, bénéficie d'un nombre de places limité. Si le jeune est sélectionné, l'ADIE se chargera de le rappeler. Cette prise de contact permet de repérer parmi les personnes intéressées lesquelles sont éligibles au programme. Les personnes n'ayant pas le profil sont réorientées vers d'autres programmes de l'ADIE. Les personnes répondant aux critères définis par l'ADIE

sont séparées en deux groupes. Un groupe n'est pas rappelé : il s'agit du groupe de contrôle. Un groupe est rappelé et incité à participer : il s'agit du groupe de traitement.

SCHÉMA 3

Le dispositif expérimental du projet CréaJeunes



Le pilotage de l'expérimentation repose sur une collaboration étroite entre les services de l'ADIE chargés de la mise en œuvre du programme et les chercheurs qui doivent assurer la supervision scientifique de l'évaluation. Le protocole est ainsi spécifié de façon précise par une concertation entre les deux parties. Les services de l'ADIE sont ensuite responsables d'assurer que celui-ci soit respecté. Les populations sur lesquelles le programme est évalué doivent notamment être représentatives de la population éligible et le « tirage aléatoire » distinguant les deux populations exposées différemment au programme effectué avec une grande rigueur. Les évaluateurs s'engagent à communiquer régulièrement sur l'état d'avancement des travaux et des résultats lors d'un Comité de pilotage, auquel participent également des représentants du Haut commissariat aux solidarités actives contre la pauvreté.

Les enquêtes

Pour évaluer le programme, le choix a été fait de s'appuyer sur les résultats de deux enquêtes, l'une un an après l'entrée dans le programme et l'autre deux ans après, les informations recueillies dans les fichiers étant frustrées et ne permettant pas en effet de donner une appréciation très riche de l'impact du programme. Bien que coûteuses en raison de l'exigence d'un taux de réponse élevé, ces enquêtes ont été jugées nécessaires. L'intérêt central des évaluations par expérimentation est en effet d'apporter une appréciation sur les effets du dispositif reposant sur le moins d'hypothèses possibles. Même si on ne saura comment traiter les individus pour lesquels on n'aura pas d'informations, faudra-t-il les ignorer en considérant que l'effet du programme est pour eux en moyenne le même que pour les individus pour lesquels on aura des informations ou faudra-t-il leur imputer une valeur à définir ?

Ces deux enquêtes permettront de mesurer l'impact du programme sur un grand nombre d'indicateurs, notamment la création d'entreprise, la survie des entreprises créées, leur accès au crédit, l'emploi et les revenus générés par ces entreprises, etc. En outre, on pourra,

à partir de ces enquêtes, procéder à une analyse coût/bénéfice du programme qui pourra notamment servir de base pour une éventuelle décision d'élargissement. Enfin, les résultats plus fins et les conclusions que l'on pourra en tirer permettront de guider les politiques publiques ayant trait à l'insertion des jeunes, au soutien à la création d'entreprise ou encore au développement du microcrédit.

Conclusion

D'une façon générale, les études d'impact actuellement réalisées en France souffrent d'un déficit de crédibilité qui empêche leurs résultats de peser sur le débat public. Cependant, évaluer de façon rigoureuse, fiable et précise est possible. Les évaluations par expérimentation permettent de surmonter nombre de problèmes dont souffrent d'autres évaluations aux hypothèses souvent restrictives. Le biais de sélection y est en effet neutralisé de façon rigoureuse et transparente. Développer ces évaluations est ainsi une bonne voie pour construire un socle de connaissances communes s'imposant à tous et sur lesquelles peut s'appuyer le débat.

Il y a en effet des programmes qui marchent, d'autres qui ne marchent pas et des programmes qui ont des effets contraires à ceux qui sont attendus. Il est nécessaire de pouvoir distinguer entre ces programmes, notamment pour réorienter l'argent public là où il est le plus efficace. Les évaluations par expérimentation sont l'outil économétrique le plus efficace pour le faire. Il s'agit d'un thème d'investigation en plein essor. Au sein de l'École d'économie de Paris, le *Poverty action lab* est une référence pour le développement de ces méthodes en France.

Les méthodes de différences de différences, de variables instrumentales et de régressions sur discontinuités

Pauline Givord (INSEE)

Quel est l'impact du suivi des chômeurs sur leurs chances de retour à l'emploi ? Les études permettent-elles d'espérer des salaires plus élevés ? Des classes à effectifs réduits améliorent-elles les résultats scolaires des élèves ? C'est à ce type de questions que tentent de répondre les méthodes qui vont être présentées ici. En général, il n'est pas simple d'y répondre directement : une corrélation observée entre deux variables ne signifie pas que l'une cause l'autre.

L'objectif des méthodes présentées ici est justement d'identifier, *i.e.* d'isoler l'effet propre d'une mesure.

Notons en préambule qu'il existe deux approches assez différentes pour traiter de ces questions : la première, « structuraliste », consiste à expliciter un modèle complet du comportement des agents concernés. Ce type d'approche est particulièrement utile *ex-ante*, pour prévoir le résultat d'une politique qui n'a jamais été appliquée auparavant. Une autre approche, plus empirique, consiste à évaluer directement l'effet d'une mesure (par l'exemple l'effet d'une formation sur l'emploi), sans chercher à expliquer les mécanismes économiques à l'œuvre : on parle d'estimation en forme réduite. C'est à ce deuxième courant que se rattache cette présentation⁶.

Inférence causale et groupe de contrôle

Le problème fondamental de l'inférence causale

Dans la suite de cet exposé, nous allons nous placer dans le cadre dit « de Rubin », du nom du statisticien qui a popularisé cette approche au cours des années soixante-dix. Son vocabulaire emprunte beaucoup au champ de l'expérimentation médicale, dont elle est issue. Ainsi, on parlera de « traitement » pour désigner la variable dont on cherche à mesurer l'effet : généralement la participation à un programme ou encore l'introduction d'une nouvelle politique ou d'une réforme.

6. On trouvera une présentation plus détaillée de l'ensemble de ces méthodes dans Givord, INSEE, 2010.

Il est loin d'être évident d'identifier l'effet propre d'une politique c'est-à-dire l'effet qui ne peut être attribué qu'à la mesure. En effet, il faudrait savoir idéalement ce qu'aurait été la situation des bénéficiaires s'ils n'avaient pas participé au programme. Cette situation, qu'on appellera contrefactuelle, ne peut être observée parce que, par définition, les bénéficiaires du programme y ont participé. Tout l'enjeu des méthodes va être d'estimer ce contrefactuel, le plus souvent en utilisant un groupe d'individus qui n'ont pas participé au programme mais dont les caractéristiques sont proches de celles des individus qui y ont participé : le groupe « de contrôle ».

Un groupe de contrôle crédible est un groupe de personnes dont on pense que les résultats auraient été proches de ceux des bénéficiaires si elles avaient participé au programme. La détermination de ce groupe de contrôle est loin d'être évidente. La première méthode qui vient immédiatement à l'esprit est de comparer les bénéficiaires et les non-bénéficiaires de la politique évaluée. Toutefois, les bénéficiaires d'une mesure ont en général des caractéristiques assez différentes des non-bénéficiaires. Aussi, en considérant simplement la différence de situation, on risque de se tromper lourdement car cette différence est très souvent due, au moins en partie, à ces caractéristiques différentes : on parle alors d'effet de sélection. On peut aussi penser à suivre l'évolution des bénéficiaires, ou plus simplement à comparer leur situation avant et après le programme. En faisant cela, on peut en fait capter beaucoup d'autres choses que l'effet du programme. Par exemple, si on considère le marché du travail, la conjoncture jouera un rôle important.

La construction du groupe de contrôle

L'enjeu des méthodes économétriques présentées ici est donc de construire ce contrefactuel, ce groupe de contrôle aux caractéristiques les plus proches possibles du groupe des bénéficiaires (ou groupe de traitement). L'idéal est bien sûr d'affecter aléatoirement une population cible au groupe de traitement ou au groupe de contrôle. Cela revient en quelque sorte à tirer au sort les bénéficiaires et à s'assurer qu'en moyenne, bénéficiaires et non-bénéficiaires ont des caractéristiques semblables. Ces méthodes ont connu un développement rapide tout récemment dans les pays anglo-saxons et en économie du développement et commencent à se généraliser en France (voir présentation précédente). Cela n'est toutefois pas toujours possible, parfois pour des raisons politiques (quand on veut par exemple que le programme soit mis en place seulement dans certaines zones), parfois pour des raisons éthiques (quand il s'agit par exemple de priver une personne dans le besoin d'une aide qui peut s'avérer vitale) ou encore parce que le protocole expérimental n'a pas été mis en place suffisamment tôt (cette méthode exige en effet de prévoir la façon dont on va évaluer la politique dès sa mise en œuvre).

À défaut, on peut utiliser des dispositifs « quasi expérimentaux » aussi appelés « expériences naturelles » par opposition aux expérimentations qui sont dites « expériences contrôlées ». On utilise dans ce cas le fait que certaines personnes aient été exclues ou partiellement exclues du bénéfice du programme évalué soit parce que les législations qui s'appliquent à elles sont différentes, soit parce que certaines de leurs caractéristiques personnelles limitent leur accès au dispositif. Dans certains cas, on peut considérer que ces différences sont dues au hasard. En règle générale, il faut faire l'hypothèse que les différences ainsi constatées, différences de législation ou de caractéristiques individuelles (appelées variables instrumentales), expliquent la non-participation ou la moins forte participation au programme évalué mais sont indépendantes des variables qui ont une incidence sur le résultat.

Ces méthodes n'impliquent pas de mobiliser des méthodes statistiques compliquées. Il faut en revanche accorder une grande importance aux hypothèses sur lesquelles elles sont fondées. Si ces hypothèses ne sont pas valides, les résultats dont elles découlent ne seront pas corrects.

Cette présentation présente les trois méthodes principales :

- les différences de différences ;
- les variables instrumentales ;
- les régressions sur discontinuités.

Les différences de différences

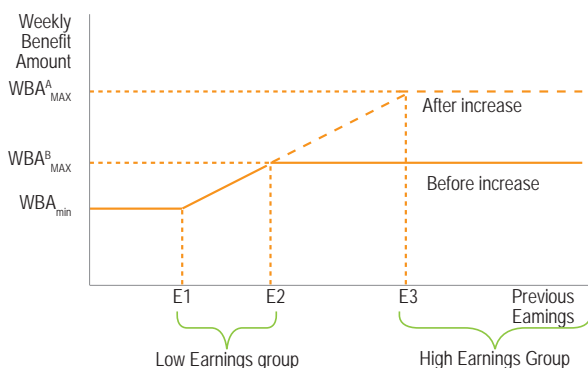
Le principe de l'estimateur dit « différence de différences » est de combiner les deux estimateurs évoqués en introduction : l'estimateur en coupe, soit la différence des situations des bénéficiaires et des non-bénéficiaires après la réforme (ou programme), et l'estimateur « avant-après », qui évalue la différence des situations des bénéficiaires avant et après la réforme.

Le premier estimateur est biaisé par l'effet de sélection mais neutralise l'effet dû à la conjoncture. Le deuxième estimateur est biaisé *a contrario* par l'effet de la conjoncture mais pas par l'effet de sélection. Une combinaison des deux peut neutraliser les deux effets : il s'agit de comparer l'évolution des situations des bénéficiaires avant et après la réforme avec celle des situations des non-bénéficiaires. En comparant bénéficiaires et non-bénéficiaires au même moment, on neutralise l'effet de la conjoncture. En considérant l'évolution, on neutralise l'effet de sélection sous l'hypothèse que celle-ci ne joue qu'en niveau sur les variables considérées. Ainsi, la démarche n'est valide que si l'on suppose que les deux groupes auraient eu les mêmes évolutions en l'absence de réforme. Évidemment, cela suppose notamment de bénéficier de données préalables à la réforme sur les bénéficiaires et les non-bénéficiaires, ce qui n'est pas toujours le cas.

C'est l'approche utilisée par Meyer, Durbin et Viscusi (1995) pour estimer l'impact du taux de remboursement sur la durée des arrêts de travail. Les auteurs utilisent une modification du plafond de revenu pour le remboursement, intervenue dans deux États américains. *A priori*, seules les personnes avec un revenu élevé sont donc concernées : les personnes dont le revenu est susceptible d'être supérieur au plafond (graphique 8). Les personnes dont les revenus sont inférieurs au plafond avant et après la réforme ne sont pas touchées par celle-ci : il s'agit du groupe de contrôle. Les personnes dont les revenus sont supérieurs au plafond avant et après la réforme reçoivent un revenu différent avant et après celle-ci : il s'agit du groupe étudié dans l'article, le groupe de traitement.

GRAPHIQUE 8

Réforme du plafond des indemnités journalières



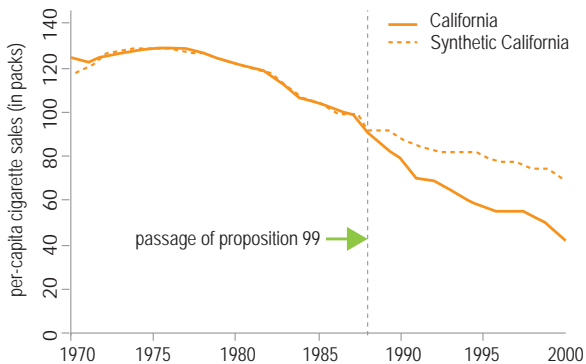
Lecture • Avant l'augmentation du plafond, les salariés gagnant E3 touchaient WBA_{Max}^B ; après l'augmentation, ils touchent WBA_{Max}^A .

Dans cet exemple, le groupe de contrôle est constitué de manière *ad hoc*. On peut penser néanmoins à d'autres façons de faire. Par exemple, pour chaque bénéficiaire, on peut chercher un non-bénéficiaire dont les caractéristiques sont aussi proches que possible de celles du bénéficiaire « jumeau » : c'est l'objet des méthodes d'appariement ou *matching*. La comparaison des « jumeaux » donne alors une estimation de l'effet du traitement sous l'hypothèse que la participation au traitement est indépendante des caractéristiques inobservables, conditionnellement aux caractéristiques observables. Cette méthode fait partie d'un ensemble plus vaste constitué de méthodes qui supposent que les données dont on dispose permettent de contrôler entièrement le biais de sélection.

Par exemple, dans le cas d'une réforme n'ayant lieu que dans certains États, on peut rechercher les individus dans des États où la réforme n'a pas eu lieu. On peut aussi raisonner directement avec des données agrégées si elles sont plus facilement disponibles. C'est la façon dont procèdent Abadie, Diamond et Hainmueller en 2007 pour estimer l'efficacité de la « proposition 99 » prise en Californie en 1988 pour diminuer la consommation de tabac (graphique 9). Les auteurs comparent directement l'évolution des ventes de paquets de cigarettes en Californie de celle dans un mix d'États supposés comparables qui n'ont pas voté cette loi. On observe une évolution très proche avant la réforme, mais les deux courbes divergent après celle-ci. En conclusion, les auteurs estiment que la loi a permis une diminution de près de 50% de la consommation de cigarettes en l'an 2000.

GRAPHIQUE 9

Consommation de cigarettes par personne en Californie par rapport à un panel d'États comparables



Lecture • En 1990, une personne résident en Californie consommait en moyenne 80 paquets de cigarettes par an.

Les variables instrumentales

Idee générale

Une autre méthode couramment retenue consiste à utiliser des variables instrumentales, c'est-à-dire des variables qui expliquent la participation à un programme mais qui ne sont pas directement corrélées avec le résultat auquel on s'intéresse. En général, on utilise des variables expliquant la participation au programme mais que l'on suppose indépendante des caractéristiques des individus, en particulier de leurs caractéristiques inobservées (on peut toujours tenir compte des caractéristiques observées par des méthodes d'appariement).

Cette méthode est fondée sur la variation de participation induite par la variable instrumentale : on s'intéresse aux personnes qui ont participé au programme parce que la variable instrumentale avait un certain niveau pour elles mais qui n'auraient pas participé (ou aurait eu une probabilité de participation plus faible) si celle-ci avait pris une autre valeur. On compare ici les bénéficiaires qui le sont seulement parce qu'ils ont été exposés à une certaine valeur de la variable instrumentale aux non-bénéficiaires qui ne participent pas uniquement parce qu'ils ont été exposés à une autre valeur de la variable instrumentale. Si la valeur de la variable instrumentale ne dépend pas des caractéristiques des individus, en particulier celles qui ont une influence sur le résultat (c'est-à-dire le paramètre d'intérêt), ces deux groupes sont comparables.

Limites

La principale limite de cette méthode tient à la réduction de l'analyse à une population bien particulière, celle pour laquelle on dispose d'une variable instrumentale permettant de scinder la population en deux catégories, la population test et la population de contrôle. L'effet estimé ne correspond pas à l'effet du traitement sur l'ensemble de la population que sous des hypothèses d'homogénéité du traitement qui peuvent être discutables. Par ailleurs, toutes les variables instrumentales ne sont pas de « bonnes » variables. Pour être utilisables, celles-ci doivent être suffisamment explicatives des différences de participation, mais en revanche non explicatives des résultats. En pratique, ces deux objectifs sont souvent antagonistes : pour trouver des variables instrumentales vraiment crédibles, c'est-à-dire indépendantes des déterminants inobservables du résultat, le risque est de choisir des variables instrumentales qui n'expliquent que très faiblement le taux de participation.

En pratique

En pratique, de « bonnes » variables instrumentales peuvent être fournies par des différences législatives. Par exemple, Lien et Evans (2005) étudient l'impact du tabagisme maternel sur le poids des enfants à la naissance en utilisant comme variable instrumentale les différences de lois anti-tabac entre les États américains. Ces lois ont en effet certainement une influence forte sur le tabagisme mais il est peu vraisemblable qu'habiter dans un État ayant une législation anti-tabac stricte ou un autre menant une politique plus laxiste soit corrélé avec les déterminants du poids de l'enfant à la naissance autres que le tabagisme de la mère. On est donc bien dans le cas d'une quasi-expérience. Les caractéristiques géographiques sont une autre source riche de quasi-expériences. Certains utilisent ainsi la distance du lieu d'habitation à certains services publics pour étudier leur efficacité. L'étude de Mc Cellan *et al.* (1994) évalue par exemple l'efficacité de certains traitements pour les accidents cardiaques en utilisant la distance aux hôpitaux mettant en œuvre ces protocoles. Les événements imprévus peuvent également être utilisés : une grève dans les transports publics dans un État américain permet à Evans et Lien (2005) de tester l'efficacité des visites prénatales. Enfin, la préférence pour la différence faisant qu'on préfère toujours avoir des enfants de sexe différent, la probabilité d'avoir un troisième enfant est plus importante si les deux premiers sont du même sexe. Ce « hasard » de la nature peut ainsi être mis à profit par exemple par Angrist et Evans (1998) pour étudier l'impact du nombre d'enfants sur l'activité féminine et pallier le fait qu'en général les femmes qui ont plus d'enfants accordent moins d'importance à leur carrière par exemple.

Les régressions sur discontinuités

Idée générale

Un troisième type de méthode est en réalité un cas particulier de la méthode des variables instrumentales : il s'agit des régressions sur discontinuités. Elles sont fondées sur le fait que de nombreux dispositifs comprennent des seuils qui permettent de limiter l'éligibilité des bénéficiaires potentiels. Il peut s'agir notamment de seuil de revenus pour la perception d'une allocation ou alors d'une condition d'âge pour accéder à un dispositif d'aide à la création d'entreprise, etc. Ainsi, la probabilité de bénéficier de ces dispositifs est discontinue, parfois fortement.

Les personnes qui sont juste au-dessous du seuil ou juste au-dessus ont généralement des caractéristiques très proches. On peut ainsi considérer que les différences d'âge ou de revenus entre ces personnes sont trop faibles pour qu'elles aient un impact significatif sur leur situation après qu'elles aient participé ou non au programme. Il en va ainsi d'une personne qui a un revenu annuel de 12 980 euros et d'une autre qui a un revenu de 13 030 euros. Ou encore d'une personne née le 29 décembre 1978 et d'une autre née le 5 janvier 1979. On peut donc les comparer sans avoir d'effets de population trop importants. Dans la fenêtre qu'on choisit, l'âge ou le revenu sont considérés comme des variables instrumentales.

En termes de mise en œuvre, ces méthodes demandent cependant d'avoir des données suffisamment précises et nombreuses pour qu'un nombre significatif d'entre elles se trouve juste au-dessus du seuil ou juste en dessous et qu'on obtienne ainsi une mesure statistique fiable. Plus précisément, il s'agit de s'assurer que l'estimateur retenu permette de mesurer des effets mêmes réduits.

Deux exemples

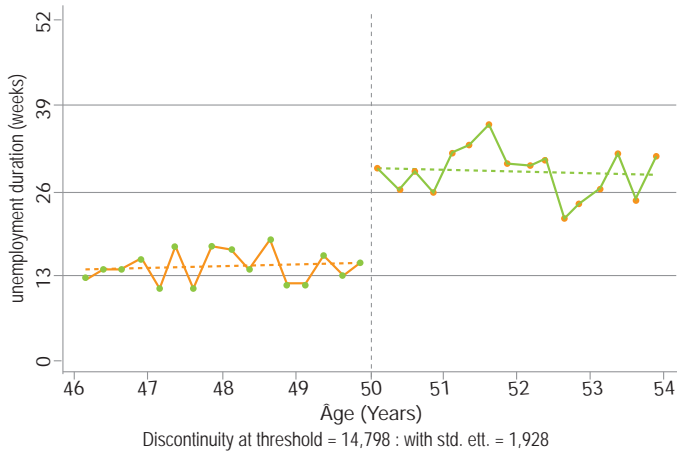
Nous allons détailler deux exemples d'application de cette méthode. Le premier reprend un papier de Laville paru en 2008 dans le *Journal of econometrics*. Ce papier présente une réforme qui a eu lieu dans certaines régions autrichiennes et qui donne une durée d'indemnisation plus longue aux chômeurs de plus de 50 ans. Le but est d'en estimer l'impact sur la durée avant le retour à l'emploi pour déterminer si la durée d'indemnisation a une influence sur la durée de chômage. Les auteurs ont simplement comparé la durée du chômage en semaine pour les personnes de moins de 50 ans et celle pour les personnes de plus de 50 ans après la réforme (graphique 10).

Les résultats sont très clairs :

- les personnes entre 46 et 49 ans ont une durée moyenne de chômage de 13 semaines, cette durée étant relativement stable en moyenne entre les personnes d'âge différent ;
- les personnes âgées de 50 à 54 ans ont également une durée de chômage stable, mais celle-ci se situe aux alentours de 38 semaines, soit une différence de près de 15 semaines.

GRAPHIQUE 10

Durée de chômage selon l'âge

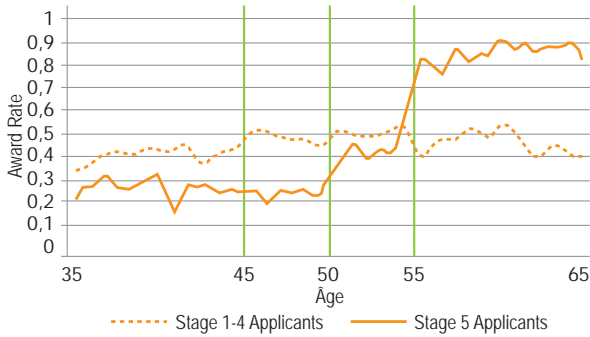


Lecture • Une personne basculant dans une période de chômage à 51 ans y reste en moyenne 26 semaines.

Le deuxième article de Chen et van der Klaauw (2008) étudie une question proche en cherchant à déterminer l'impact des allocations d'invalidité aux États-Unis sur l'inactivité. Les auteurs utilisent pour ce faire le mécanisme en vigueur aux États-Unis pour être déclaré éligible à ces allocations. Pour certaines catégories de handicap, l'éligibilité est en effet basée sur une appréciation des capacités de l'individu à s'insérer sur le marché du travail. Il existe une grille pour cette évaluation avec un seuil d'éligibilité lié à l'âge (graphique 11). Plus on est âgé, plus on a ainsi de chances d'être éligible. Le taux d'éligibilité pour la catégorie de handicaps les plus graves (classe 5) est ainsi de 25 % en moyenne si l'individu a entre 45 et 50 ans, 40 % s'il a entre 50 et 55 ans et 85 % au-dessus de 65 ans. Les auteurs choisissent alors d'étudier les personnes ayant un âge proche de 50 ans et celles ayant un âge proche de 55 ans car, quand on observe la distribution des taux d'éligibilité selon l'âge, les discontinuités sont particulièrement fortes autour de ces seuils. Les résultats sur l'offre de travail sont cependant modestes, puisque Chen et Van der Klaauw estiment que les bénéficiaires de ces allocations auraient eu une offre de travail de 20 % supérieures au plus, s'ils n'avaient bénéficié d'aucune aide. L'effet est encore plus modeste pour les personnes dont les motifs d'invalidité sont uniquement liés aux difficultés d'insertion sur le marché du travail.

GRAPHIQUE 11

Proportion d'éligibles selon l'âge



Lecture • Une personne âgée de 50 ans a 30% de chances de se voir accorder une invalidité de niveau 5 et 48% de chances de se voir accorder une invalidité de niveau 1 à 4.

Conclusion

En conclusion, ces nouvelles méthodes offrent un cadre d'analyse rigoureux reposant sur des mécanismes assez convaincants notamment par leur simplicité. Elles sont souvent assez intuitives et produisent des résultats fiables et précis. Les résultats de ces méthodes semblent cependant un peu bruts car celles-ci n'apportent pas, bien souvent, une compréhension fine des mécanismes économiques auxquels les résultats sont dus. Les résultats obtenus ne sont le plus souvent valables que pour une population particulière et dépendent fortement du contexte économique et social. En cas d'hétérogénéité des effets, dans l'espace, le temps ou entre sous-groupes aux caractéristiques différentes, l'information disponible est donc en définitive assez limitée. La possibilité de généraliser les résultats doit donc être examinée au cas par cas et reposer sur des arguments suffisamment convaincants. Si ces méthodes assurent en général la validité interne de l'évaluation, sa validité externe peut ainsi être parfois faible.

Pour aller plus loin

Aujourd'hui, l'un des grands enjeux de l'évaluation de politique publique est d'aller plus loin qu'un raisonnement en moyenne. Généralement, les méthodes économétriques ne permettent pas d'évaluer l'effet pour une personne particulière, mais se contentent d'étudier l'effet moyen sur une population donnée, plus ou moins large. Or, les chercheurs en sciences sociales et les responsables politiques aimeraient souvent en savoir plus. En effet, il existe des cas où les effets sont fortement hétérogènes. Une mesure peut avoir des effets nuls en moyenne ou presque, mais être « socialement » utile. Généralement, ce sont des politiques qui favorisent une population au détriment d'une autre pour établir un équilibre que le marché a perturbé pour assurer une plus grande justice sociale. C'est par exemple le cas des politiques redistributives ou d'autres politiques ayant pour but d'éviter l'exclusion de certaines populations.

L'étude de l'hétérogénéité de l'effet d'une réforme ou d'un programme a souvent été laissée aux chercheurs en sociologie qui s'appuient sur des entretiens approfondis et des études de cas. On assiste cependant aujourd'hui au développement de techniques économétriques qui permettent d'évaluer l'impact des programmes non seulement en moyenne mais aussi en différence selon les catégories de la population concernée. Il s'agit notamment de régressions par quantiles.

BIBLIOGRAPHIE

- Abadie A., Diamond A., Hainmuller J.**, 2007, *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program*, NBER Technical Working Papers 0335.
- Angrist J., Evans W.**, 1998, « Children and Their Parents' Labor: Evidence from Exogenous Variation in Family Size », *The American Economic Review*, vol. 88, n° 3, p. 450-477.
- Angrist J., Krueger A.**, 2001, « Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments », *Journal of Economic Perspectives*, vol. 15, n° 4.
- Behaghel L., Crépon B., Gurgand M.**, 2009, *Évaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise, Rapport Final*, École d'Économie de Paris.
- Caliendo M., Kopeinig S.**, 2005, *Some Practical Guidance for the Implementation of Propensity Score Matching*, IZA Discussion Papers 1588.
- Chen S., Van der Klaauw W.**, 2008, « The work disincentive effects of the disability insurance program in the 1990s », *Journal of Econometrics*, vol. 142(2), p. 757-784.
- Evans W., Lien D.**, 2005, « The benefits of prenatal care: evidence from the PAT bus strike », *Journal of Econometrics*, vol. 125(1-2), p. 207-239.
- Givord P.**, 2010, *Méthodes économétriques pour l'évaluation de politiques publiques*, document de travail du D3E, INSEE, G2010/08.
- Heckman J., Vytlacil E.**, 2005, *Structural Equations, Treatment Effects and Econometric Policy Evaluation*, NBER Working Papers 11259.
- Heckman J., Ichimura H., Todd P.**, 1998, « Matching as an Econometric Evaluation Estimator », *Review of Economic Studies*, vol. 65(2), p. 261-94.
- Imbens G.**, 2000, *Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score*, Econometric Society World Congress 2000 Contributed Papers 1166, Econometric Society.
- Lalive R.**, 2008, « How do extended benefits affect unemployment duration A regression discontinuity approach », *Journal of Econometrics*, vol. 142(2), p. 785-806.
- Lien D., Evans W.**, 2005, « Estimating the Impact of Large Cigarette Tax Hikes: The Case of Maternal Smoking and Infant Birth Weight », *The Journal of Human Resources*, vol. 40, n° 2, p. 373-392.
- McClellan M., Mcneil B. J., Newhouse J. P.**, 1994, « Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables », *The Journal of American Medical Association*, vol. 272(11), p. 859-866.
- Meyer B. D., Durbin D. L., Viscusi W. K.**, 1995, « Workers' Compensation and Injury Duration: Evidence from a Natural Experiment », *American Economic Review*, vol. 85(3), p. 322-340.

ÉVALUATION ET CHANGEMENTS ORGANISATIONNELS

L'apport des études de terrain

La troisième séance du séminaire sur les méthodes d'évaluation de politiques publiques était consacrée aux méthodes qualitatives de l'évaluation et, plus particulièrement aux interactions existant entre les démarches évaluatives et les démarches d'actions des acteurs ou des entités auxquelles elles se rapportent.

– L'intervention de Jean-Claude Moisdon, sociologue et ancien directeur du Centre de gestion scientifique de l'École des mines de Paris, était centrée sur les méthodes de recherche-intervention qui permettent d'étudier à l'aide d'un processus interactif les changements organisationnels induits par une politique publique.

– Hélène Genin, consultante, directrice des politiques Santé et Solidarités au cabinet Euréval, a, quant à elle, retenu l'approche monographique pour analyser les actions entreprises au niveau local par les acteurs institutionnels pour répondre aux objectifs des politiques publiques dont ils ont la charge.

Ces deux interventions ont été suivies d'un débat avec la salle.

L'approche par la recherche-intervention

Jean-Claude Moisdon (ENSMP)

Avec quelques collègues chercheurs en gestion, je pratique une méthode qui porte le nom de recherche-intervention. Elle s'appuie sur l'idée qu'il n'est possible de comprendre vraiment le fonctionnement d'une entreprise qu'en y pénétrant, en y intervenant et, par conséquent, en la modifiant. Nous sommes par là même *a priori* assez loin de l'évaluation qui semble supposer un écart entre l'analyste et son objet pour peser de la façon la plus neutre possible sur les résultats d'une action ou d'une politique. Pourtant sur le thème particulier de l'évaluation du changement organisationnel, notre conviction est que cette dernière n'est possible que si elle s'accompagne d'une certaine forme d'interaction, qui constitue le cœur de la méthode de recherche-intervention.

C'est dans cette perspective générale que nous avons été amenés à participer à une évaluation réalisée pour la DREES concernant l'influence de la tarification à l'activité (T2A) sur l'organisation hospitalière.

Je reviendrai dans un premier temps sur les questions de méthode, que j'illustrerai ensuite par un résumé des observations que nous avons pu faire lors de cette opération.

Recherche-intervention et évaluation

Des questions difficiles

Un certain nombre de problèmes se posent lorsque l'on veut essayer de saisir, dans un système social quelconque (entreprise industrielle ou de service, administration, secteur, etc.) les liens entre une action donnée ou un ensemble d'actions et les changements organisationnels.

Tout d'abord existent différents types de changements. Les changements d'organigramme, de répartition des activités et des fonctions, sont les plus décisifs pour l'avenir de l'entité analysée, et les plus intéressants à étudier. On sait depuis longtemps que d'autres évolutions sont à considérer ; on peut évoquer les règles de coordination, la culture, les relations ou encore les pratiques. Dans ces conditions, il faut évidemment se départir de l'illusion qu'existe un arbitre suprême capable de conduire seul le changement. En fait, celui-ci fait intervenir des processus complexes et une multitude d'acteurs.

Par ailleurs, une fois entrevus les types de changement, il est nécessaire dans l'évaluation de se pencher sur leurs conséquences, plus ou moins étendues, mais aussi sur leurs causes. Ces dernières peuvent être internes ou externes. Elles sont la plupart du temps multiples, et se posent à la fois comme ailleurs, mais sans doute de façon particulièrement aiguë, la question de la notion même de causalité. Toutes ces problématiques donnent lieu à d'importants débats et polémiques dans les rangs des gestionnaires.

Organisation, dispositifs et outils

Ces problèmes viennent pour l'essentiel du fait que le terme d'« organisation » s'avère lui-même extrêmement polysémique et ambigu. Il qualifie soit l'état des rapports de coordination, de coopération et de pouvoir à un moment donné, soit l'acte d'organiser. Il s'agit là de deux conceptions fort différentes. Dans le cadre de la première conception, l'organisation peut se traduire par un découpage particulier en diverses unités, des modes hiérarchiques, des processus et des instruments de pilotage. De même, les relations interindividuelles, les phénomènes de pouvoir, les relations entre les groupes ou encore les relations professionnelles entre syndicats et patronat doivent être pris en compte. Une organisation se caractérise aussi par une culture interne, c'est-à-dire par la façon dont elle se fabrique un système de représentations et de paradigmes propre, et une culture sociétale, car elle est traversée par des phénomènes culturels venant de l'extérieur. Un certain nombre d'excellentes synthèses en la matière (Desreumaux, 2005) montrent à l'envi ce caractère polymorphe.

Ces ambiguïtés sur le sens même du terme et la multiplicité des angles d'attaque possibles posent des problèmes de visibilité et d'objectivation. Il faut alors éviter le piège qui consiste à prendre pour argent comptant les discours des acteurs, dont les perceptions de l'ensemble sont souvent autant limitées que celles de l'observateur. Au sein des hôpitaux par exemple, beaucoup déclarent spontanément que la communication et le dialogue de gestion se sont améliorés avec la mise en place de la T2A. Or, des observations sur longue durée montreraient que les relations entre médecins et administratifs se sont transformées dès la mise en place du budget global dans les années 1980 et que l'opposition systématique, classiquement évoquée auparavant, a laissé la place à des rapprochements et même des connivences significatives face au régulateur. Non seulement il n'est en rien évident de spécifier laquelle de ces deux temporalités est juste, mais on peut se poser la question des conséquences tangibles d'une telle évolution, au-delà d'une pacification de rapports jugés jusque-là hostiles.

En fait, à nos yeux, il convient de quitter rapidement l'ordre du discours pour aller vers le construit. L'organisation peut être en première approximation considérée comme une collection d'artefacts, tant il est vrai que, depuis le milieu du XIX^e siècle, elle se structure à partir de deux objets essentiels, à savoir les dispositifs et les outils (Moisdon, 1997). De tout temps, la conduite des affaires s'est appuyée sur des dispositifs. Ceux-ci sont des arrangements construits et visibles des acteurs et des objets dans l'espace et le temps. Il s'agit donc de règles sur le temps de travail, sur la nature des tâches, de plannings de réunions, de délais, etc. S'ils ne sont pas nouveaux (ils constituent par essence l'action organisée), ils deviennent de plus en plus sophistiqués. En revanche, les outils sont apparus bien plus récemment. Ils représentent des formalisations, s'appuyant souvent sur l'édifice comptable, telles que les indicateurs, les tableaux de bord, les analyses statistiques, les modèles d'aide à la décision, etc. Ils mettent en rapport diverses variables (des quantités produites, des prix, des effectifs, des signaux de qualité, etc.) et en quelque sorte formalisent la relation entre la gestion de l'entité considérée et la performance. Tous les univers de l'activité productive (de l'industrie à la culture) se sont retrouvés rapidement envahis par ces outils de gestion.

D'une façon générale, les outils et les dispositifs permettent de :

- définir et déléguer les activités ;
- coordonner les activités ;
- évaluer les activités.

On peut considérer, en effet, qu'en première approximation l'action collective organisée se traduit tout d'abord par une répartition des activités et des rôles, qu'il faut ensuite relier entre eux par des outils et des dispositifs de coordination. En troisième lieu, afin de vérifier si l'ensemble fonctionne correctement, une évaluation est entreprise en s'appuyant, en général, sur des outils permettant la hiérarchisation et le jugement.

Ne pas se contenter des discours signifie qu'il faut d'abord s'attacher à ces deux entités dont de nombreuses observations ont montré qu'elles étaient profondément structurantes (Berry, 1983). Un changement de culture ou de relations peut être déterminant mais, alors, il s'accompagne toujours d'évolutions corrélatives dans les outils et dispositifs qui sont des médiateurs entre les acteurs et la performance et qui demeurent donc centraux dans l'analyse ; d'autant qu'une fois mis en place, ils pèseront fortement sur les actions individuelles et collectives, les orientant dans des directions parfois inattendues.

Évidemment, cette position en quelque sorte « technologique » n'épuise pas le sujet des déterminants de l'action collective. Elle est plutôt un principe de méthode qui consiste à aller d'abord vers le visible, en faisant le pari qu'en retour le travail sur les techniques gestionnaires permettra de rendre visible ce qu'elles ne laissent pas apparaître spontanément.

Le changement organisationnel

Pour autant, si l'on se contente dans un premier temps des outils et dispositifs, cela ne saurait suffire. Dans quelle mesure jouent-ils vraiment sur les fonctionnements en cause ? Il faut distinguer, en effet, quatre niveaux d'appropriation des dispositifs et des outils (Grimand, de Vaujany, 2006) :

- l'appropriation cognitive (on connaît) ;
- l'appropriation formelle (on fabrique) ;
- l'appropriation discursive (on échange à partir de...);
- l'appropriation réflexive (on change grâce à...).

C'est à ce dernier niveau que l'organisation commence vraiment à évoluer. L'outil et le dispositif se trouvent alors imbriqués étroitement avec l'action. Dans les autres niveaux, il n'y a qu'apparence de changement : la création instrumentale n'embraye pas réellement sur le cours de l'action.

Le problème de l'évaluation réside, alors, dans la confusion entre les niveaux : la construction d'un outil ne suffit pas à induire un changement. L'enquête est indispensable pour déterminer ce qui a été produit de nouveau quant au système d'outils et de dispositifs, mais elle doit être complétée par d'autres investigations. En effet, au niveau des organisations se mêlent croyances et simulacres sur les effets réels des artefacts que les uns ou les autres créent pour réguler l'action collective. Souvent, on croit en l'effet qui de fait est absent, car on est loin du théâtre des opérations. D'autres fois, on fait semblant d'adopter des prescriptions gestionnaires qui, n'étant pas de l'ordre du juridique, permettent un grand nombre de tactiques diverses.

Quelle est alors la nature de ces autres investigations nécessaires ?

La méthode de recherche-intervention

D'après Hatchuel (2000) toutes les méthodologies de construction des faits scientifiques peuvent se réduire à trois postures, celle du laboratoire, celle du terrain et celle de la recherche-intervention.

Au sein du laboratoire, les objets sont manipulables et confinés. Cependant, parfois, il n'est pas possible de manipuler l'objet en laboratoire. Dès lors, les études doivent être menées sur le terrain, géologique par exemple. L'objet ne peut être confiné ni manipulé, mais il peut être naturalisé, en ce sens qu'il n'est pas transformé par le regard de l'observateur. Enfin, la recherche-intervention s'applique aux objets qui ne peuvent être ni confinés, ni manipulés, ni même naturalisés. Dans ce cadre, dès lors que le chercheur l'observe, l'objet change. Il s'agit là d'une réalité qu'il se doit d'accepter.

Une autre façon de différencier les approches consiste à poser que dans le champ des sciences sociales, le prélèvement s'oppose à l'interaction. Le prélèvement, qui est de l'ordre du terrain (les mathématiques de la décision, par exemple, étant de l'ordre du laboratoire) consiste à chercher des données sur l'entité analysée et à les exploiter sans interférer avec cette dernière. L'interaction suppose par essence que les chercheurs entrent en dialogue avec les acteurs et discutent de leur vision de la situation et de leurs projets, qui, de façon variable mais systématique, se trouvent modifiés par cette rencontre.

Dans le domaine de la gestion plus particulièrement, cette interaction peut se faire et doit se faire à nos yeux autour des outils et dispositifs. Il ne s'agit donc pas, d'emblée, de modifier par exemple la culture en place mais d'accompagner une organisation qui s'est donnée pour but de changer en travaillant avec elle sur les règles et instruments concrets qu'elle compte développer. Cette posture permet, en effet, de recueillir beaucoup plus d'informations sur la réalité du changement que des entretiens (qui restent évidemment nécessaires). C'est par les réactions des acteurs à la mise en scène d'un scénario instrumental concret que le chercheur intervenant peut saisir les leviers d'une action collective renouvelée ou au contraire les obstacles susceptibles de s'y opposer. C'est ainsi, dans une sorte de boucle de retour, que sont repérés les instruments de gestion mis en place, « déjà là », éventuellement discordants par rapport à ceux que l'on souhaite implanter, mais aussi les pratiques indigènes, non régulées au sommet, par lesquelles l'action organisée se structure (Reynaud, 1998).

Les champs de force

Dans le cadre de l'interaction et de la recherche intervention, il faut savoir assumer le biais observationnel, loin de la logique des groupes témoins. Aucun test statistique ne permet d'obtenir la « vérité ». Les conclusions sont en fait fondées sur l'intime conviction du chercheur, qui se retrouve dans l'arène pour éprouver lui-même les champs de force qui traversent l'organisation et qui la mettent en tension. La pression financière, les conditions de travail, les relations avec l'environnement, l'efficacité des processus internes, la qualité des produits, etc., tous ces termes constituent l'expression d'autant de champs de force qui tirent l'action collective dans des directions différentes. À un instant donné certains dominent les autres, selon les organisations et leurs trajectoires, et cette domination confère, au moins pour un temps, une stabilité.

Mais des déséquilibres s'introduisent fatalement un jour ou l'autre, à un stade donné de l'histoire de l'organisation, et il faut alors distinguer les forces actives des forces réactives. Les premières sont celles que les acteurs souhaitent privilégier, alors que ce n'était pas le cas, et c'est à partir d'elles qu'ils construisent de nouveaux dispositifs et outils. Les forces réactives tirent les acteurs dans un sens qu'ils ne souhaitent pas forcément et opposent aux forces

actives leurs systèmes propres d'évaluation, tout comme les routines de délégation et de coordination qui leur sont associées. Le chercheur va donc essayer de décrire une rencontre entre une organisation émergente et une organisation existante, toujours résistante.

Finalement, l'intervention se recoupe très largement avec l'évaluation, en ce sens qu'elle a pour but premier de reconstituer des déséquilibres dans les champs de force et de dessiner des issues possibles vers un rééquilibrage provisoire.

Tarification à l'activité et organisation hospitalière

L'évaluation menée autour de l'impact sur l'organisation de la tarification à l'activité s'inscrit dans le cadre du dispositif général d'évaluation de la T2A piloté par la DREES. Au passage, il est remarquable que l'État s'intéresse à la question des organisations, au-delà du seul aspect des résultats, c'est-à-dire à une intermédiation essentielle entre une politique publique (on fera l'hypothèse que le modèle de financement et ce qui l'accompagne en est une) et ses impacts.

L'organisation hospitalière

Chacun sait que l'organisation hospitalière est très particulière (Mintzberg, 1982 ; Glouberman *et al.*, 2001). L'établissement de santé, dans son ensemble, vit des évolutions rapides des pratiques médicales et des équipements, mais l'organisation, elle-même, qui répond à une logique de balkanisation associée à une spécialisation et une technicisation croissantes, ne change en réalité, en quelque sorte, que quantitativement, au travers de l'accentuation continue de ces caractéristiques. Le nombre de services augmente malgré la création récente des pôles⁷, pour le secteur public, ainsi que celui des équipements sophistiqués.

Si l'on reprend les trois fonctions des dispositifs-outils, les mécanismes de délégation s'avèrent de plus en plus spécialisés. Les mécanismes de coordination restent faibles et les mécanismes d'évaluation apparaissent disjoints. Les administratifs par exemple sont aujourd'hui théoriquement évalués sur des critères économiques (alors qu'ils étaient autrefois jugés sur l'absence de tensions au sein de la structure). Les médecins sont évalués sur tout autre chose, qui relève de leur logique professionnelle (satisfaction de leurs patients, progression de leurs savoirs, développement de leur carrière, etc.).

Depuis la fin de la Seconde Guerre mondiale, les outils de gestion sont demeurés quasi inexistants, et cela jusqu'à la mise en place de la T2A. En effet, seuls quatre outils, « les quatre vieux », permettaient de juger du fonctionnement d'un hôpital (le nombre d'admissions, le nombre de journées, le taux d'occupation et la durée moyenne de séjour). La tutelle, mais également les hospitaliers eux-mêmes ne disposaient de fait d'aucune autre vision sur l'activité d'un établissement que celle fournie par ces quatre indicateurs, qui ne distinguent pas les hospitalisations pour une greffe de foie et celles relatives à une appendicite.

Le budget global a remplacé le tarif par journée, en 1983. Il s'est avéré que le premier n'incitait pas davantage à l'efficacité que le second. Au contraire, il s'est agi d'une chape de plomb qui se traduisait par une très grande opacité et par la préservation des rentes de situation, pour certains, et au contraire par celle d'une situation précaire, pour les autres.

7. Les pôles sont des regroupements de services coordonnés par un triumvirat médecin/cadre soignant/administratif. Ils ont été créés par ordonnance en 2005, et sont réservés au secteur public.

Le programme de médicalisation des systèmes d'information et la régulation « souple »

Un premier changement intervient au milieu des années 1980, avec la mise en place du PMSI (programme de médicalisation des systèmes d'information). Ce système, importé des États-Unis, permet à cette date de catégoriser les patients en 500 ensembles réputés homogènes à la fois d'un point de vue médical et d'un point de vue économique (les GHM : groupes homogènes de malades). Une étude renouvelée annuellement (l'Étude nationale des coûts) permet par ailleurs d'estimer, sur la base d'un échantillon d'hôpitaux, les coûts moyens d'un séjour dans chaque groupe homogène de malades.

L'activité n'est donc plus décrite par le simple nombre d'admissions mais par 500 nombres d'admissions différents, chacun se voyant affecter un coût moyen par patient. Dans cette logique, on peut calculer le budget théorique d'un établissement, si l'on dispose de son « *case-mix* », c'est-à-dire de son nombre d'admissions dans chaque GHM, et cela à partir de 1996. Le fait que le budget réel soit supérieur au budget théorique est censé signaler un problème d'efficience. Les agences régionales de l'hospitalisation, qui voient le jour à cette période, sont censées pénaliser les hôpitaux qui se trouvent dans cette situation et les inciter à des réorganisations. Cela dit, cette innovation dans le système régulateur n'a pas été couronnée de succès (Engel *et al.*, 2001) alors que, malgré ses imperfections nombreuses, il permettait de mettre en évidence des inégalités flagrantes sur le territoire. Ainsi, à activité identique, un hôpital pouvait se montrer trois fois moins efficace qu'un autre et par conséquent, disposer de trois fois plus de moyens.

Les causes de cet échec sont plurielles, comme d'habitude, mais l'une d'entre elles, essentielle, était que cet indicateur global (rapport entre budget théorique et budget réel), ne fournissait aucun moyen pour les hôpitaux de faire le lien avec le fonctionnement concret de l'établissement. Cet outil avait été adopté sur la base des théories économiques marginalistes (concurrence par comparaison notamment), à l'époque très en vogue. Il visait à envoyer aux producteurs de soins un signal global, en fait tellement global qu'il ne leur permettait pas de déceler les points faibles de leur organisation et restait quelque peu mystérieux.

La tarification à l'activité (T2A) et son évaluation

C'est en 2004 que la tarification à l'activité voit le jour. Un prix est fixé pour chaque groupe homogène de malades, dont le nombre ne cesse d'augmenter, pour atteindre 2 300 avec la version 11 du PMSI, mise en œuvre en 2009. C'est donc un système extrêmement complexe. Il s'appuie notamment sur des règles de codage, des catalogues de diagnostics (15 000 environ déclinés selon la nomenclature de l'OMS) et d'actes (8 000), des procédures multiples de financement (forfait au séjour, supplément à la journée pour certains types d'admission, décote du prix pour des durées de séjour atypiques, remboursement à l'euro de certains médicaments, etc.). Pour autant, le principe est clair. Les médecins le comprennent mieux que l'indicateur global précédent. Ils savent qu'ils obtiennent des recettes à l'aune de leur activité et ce système semble, en conséquence, plus incitatif. Il a aussi le mérite d'être plus proche de l'activité par essence que le précédent système. On en espère donc, non seulement un surcroît d'équité entre les établissements (rémunérés selon leur activité) mais aussi une plus grande rigueur dans la gestion, portée par des réorganisations internes.

C'est sur ce dernier thème des évolutions organisationnelles qu'une évaluation a été menée entre 2007 et 2009, par une équipe mixte de consultants (ESSOR) et de chercheurs (CGS École des mines) et sur la base d'un petit groupe d'établissements volontaires (un CHU, trois CH, deux PSPH et deux cliniques). Un échantillon si restreint et si peu représentatif peut conduire à émettre des doutes sur la véracité des conclusions. Toutefois, cette faiblesse était

compensée par l'utilisation d'autres expériences, telles que le chantier lancé par la Mission d'expertise et de l'audit hospitalier (MEAH⁹) sur la nouvelle gouvernance, ou encore les stages des élèves directeurs de l'École des hautes études en santé publique, l'EHESP (une quinzaine d'entre eux a consacré un mémoire spécial au thème des impacts organisationnels de la T2A). De plus, un travail de documentation a été mené en parallèle sur les expériences étrangères, notamment américaine. En outre, il faut savoir que le secteur hospitalier constitue un milieu extrêmement mimétique. Ainsi, une évolution se répète d'établissement en établissement. Enfin, le fait que les établissements de notre échantillon étaient volontaires, laissait entendre que ces établissements étaient porteurs de projets de changement et il devenait dès lors intéressant de voir s'ils parvenaient ou non à les concrétiser réellement. C'est ainsi que peuvent être identifiées les forces réactives (cf. ci-dessus), plutôt qu'en étudiant un hôpital passif, n'imaginant pas de transformation.

L'enquête s'est tout d'abord déclinée sous la forme d'entretiens et d'observations entre septembre 2007 et mars 2008 (ce que l'on peut appeler la phase « terrain »). Des axes de progrès potentiels ont ainsi été identifiés, comme la contractualisation entre direction et pôles, ou l'élaboration et l'utilisation des tableaux de bord médico-économiques, ou encore l'intéressement des médecins, à partir des demandes des établissements. Nous sommes, alors, entrés dans une phase d'intervention portant sur les projets ainsi identifiés, qui a duré quatorze mois et qui s'est principalement déroulée sous forme d'une participation à des groupes de travail, dans chacun des établissements.

Les effets organisationnels de la T2A

La phase « terrain » permet de constater un certain nombre de changements au niveau que nous privilégions, celui des dispositifs et des outils.

Parmi les principaux, on peut noter :

- a) des réformes profondes de la chaîne administrative liée au patient (de l'accueil à la sortie), avec pour ambition première de gérer au mieux la facturation, le but étant de récupérer aussi rapidement que possible le financement de l'assurance-maladie ;
- b) autre fait massif : de nombreux établissements ont développé des outils médicoéconomiques. L'outil CREA (compte de résultat analytique) par exemple est un instrument issu de la comptabilité analytique qui permet, pour chaque entité clinique (pôle, service, département...), de mettre en regard ses dépenses et ses recettes (générées par son case-mix particulier). C'est ainsi qu'il est possible de repérer les pôles bénéficiaires et les pôles déficitaires, de la même façon que la tutelle repère les hôpitaux bénéficiaires et les hôpitaux déficitaires ;
- c) au niveau des ressources humaines, apparaît une légère distorsion en faveur de la technostructure (recrutement de contrôleurs de gestion par exemple) ;
- d) la contractualisation interne se développe et se sophistique à la fois au niveau des règles et des indicateurs de résultat ;
- e) par ailleurs des infléchissements notables sont observables quant à la stratégie de l'établissement, qui visent de plus en plus à désenclaver ce dernier (partenariats, fusions, coopérations sur des équipements, etc.).

En revanche, la phase de travail en commun avec les acteurs hospitaliers (médecins, administratifs, cadres infirmiers...) a révélé de grandes difficultés à évoluer dans certaines direc-

9. Mission créée en 2003 et regroupée, depuis cette année, avec deux autres entités du système régulateur pour constituer l'ANAP (Agence nationale d'appui à la performance des établissements de santé et médico-sociaux).

tions, pourtant unanimement souhaitées et *a priori* cohérentes avec le nouveau mode de financement.

Deux constatations peuvent être faites à ce titre : tout d'abord, si les outils de type CREA sont pléthoriques, rares sont ceux qui s'en servent véritablement, sauf dans certaines situations extrêmes où ils permettent d'alimenter un argumentaire vis-à-vis de pôles ou de services en déséquilibre fort (et déjà connu) entre ressources et activités. De même, on se rend compte que les documents décrivant les contrats deviennent particulièrement épais et fournis en outils d'évaluation, mais que les résultats en termes d'amélioration de l'efficacité, ou aussi de la qualité, sont lents à se concrétiser.

Par ailleurs, si l'on revient au plus près de l'activité, on constate que très peu d'avancées sont faites quant à l'analyse fine des parcours des malades, de l'organisation du travail des professionnels, des processus de soins, malgré la conviction des acteurs (et des chercheurs) que c'est à ce niveau que se situent les enjeux les plus importants en termes de performance.

Au final, les évolutions internes apparaissent peu spectaculaires, mais une sortie par le haut est souvent programmée, au travers d'une alliance avec d'autres établissements. Cette même conclusion a été relevée aux États-Unis, où la mise en place d'un dispositif analogue (le PPS, *prospective payment system*, mis en œuvre dès 1983) a conduit à une vaste réorganisation au niveau des institutions (fusions, absorptions, réseaux, etc.) sans grand changement visible sur les processus de soins eux-mêmes (Bazzoli *et al.*, 2004). Contrairement aux attentes et visées du régulateur qui pensait qu'une meilleure coordination serait induite à partir d'un système de financement très détaillé, proche de l'activité, pour surmonter la balkanisation actuelle, l'organisation du travail (horaires, effectifs, modes de coordination, gestion du parcours du patient, etc.) n'a que peu évolué.

La T2A n'atteint pas le cœur de l'activité

Au final, il semblerait que la tarification à l'activité n'atteigne pas le cœur de l'activité. Cet incitatif économique est à l'origine d'un « stress gestionnaire » qui s'avère d'un apport limité dans la résolution des problèmes organisationnels structurels du secteur.

Quelles sont les causes d'un tel phénomène ?

Existente tout d'abord des difficultés et des problèmes qui ressortent du système de financement lui-même. Les professionnels, même animés de la meilleure volonté, finissent par se décourager devant la complexité des outils comme le CREA et l'instabilité des informations qui permettent de le construire, et aussi d'en comprendre les résultats. La tarification à l'activité, qui se fonde sur des milliers de règles, est en effet en perpétuelle évolution, la plupart du temps d'ailleurs pour des raisons parfaitement légitimes (amélioration des classifications médicales, innovations technologiques à prendre en charge au niveau du financement, correction du tarif de telle ou telle activité manifestement mal rémunérée dans le dispositif actuel, etc.), état de fait qui ne facilite pas l'usage interne des instruments qui en sont dérivés.

Une raison plus déterminante (et qui continuerait à peser sur le cours des choses si le système de financement se stabilisait) nous semble devoir résider dans ce que l'on peut appeler l'organisation spontanée d'un établissement de santé, quel qu'il soit (hôpital public, ou participant au service public, ou clinique, etc.), c'est-à-dire l'organisation qui résulterait, sans volontarisme transformateur particulier, de la nature de l'activité et des caractéristiques communes des opérateurs qui la prennent en charge. Cette organisation est considérée par les spécialistes de la gestion comme la plus complexe qui soit, car, d'une part, elle est hautement professionnalisée, d'autre part, elle exige, pour fonctionner sans trop de drames, un haut niveau de coordination, dans l'acceptation la plus simple et la plus technique de ce terme,

à savoir, la gestion des rendez-vous entre les différentes entités participant à la production des soins (Moisson, 2008).

Reprenons ces deux points (professionnalisation et coordination) :

- a) au niveau de l'activité de base (la séquence de soins et de prestations à mettre en place pour chaque patient, pris un à un) l'asymétrie d'information entre le management et les professionnels est importante, compte tenu de la nature des savoirs en cause ;
- b) la balkanisation inhérente à cette professionnalisation joue à l'encontre de moyens de coordination simples.

Ainsi, les trajectoires des malades, qui sont décidées par les professionnels eux-mêmes, mobilisent souvent plusieurs dizaines d'acteurs, aux métiers et aux enjeux très différents. Ceux-ci sont focalisés sur le patient et sur la prestation qu'ils lui offrent, et non sur la trajectoire d'ensemble de ce dernier. Le malade et son état restent au centre de leur activité et il est impossible de le leur reprocher, car toute leur formation et leur légitimité reposent sur cette focalisation. Cette dernière caractéristique distingue fortement l'établissement de santé d'autres organisations professionnelles (comme la recherche), où les besoins de coordination ne sont pas aussi intenses.

Mais cela veut dire qu'au total, dans cet atelier très particulier que constitue l'établissement de santé, il n'existe ni service de méthodes ni service d'ordonnancement, si l'on cherche la comparaison avec un système de production industriel. Or, un centre hospitalier moyen gère chaque année plusieurs dizaines de milliers de trajectoires qui sont, par nature, incertaines. Il convient d'ajouter que les résultats en termes de santé restent peu mesurables, malgré un certain nombre d'avancées en la matière, et que les référentiels liant le niveau d'activité à celui des ressources (humaines notamment) sont encore rares, ce qui limite la possibilité de types de gestion associés aux systèmes complexes (par les résultats par exemple).

Concrètement, l'organisation s'appuie sur des règles professionnelles, des conventions plus ou moins partagées et beaucoup d'ajustement mutuel, qui pousse les acteurs à continuellement négocier entre eux, et sur n'importe quoi : les délais, les ordres de passage, les prescriptions elles-mêmes, etc. À ce niveau de l'organisation, il n'est même plus possible de parler d'asymétrie de l'information, dans la mesure où les médecins n'en savent pas plus que les administratifs sur la façon de configurer les processus en cause afin d'assurer à la fois qualité, efficacité, conditions de travail convenables (car l'ajustement mutuel, à qui l'on peut conférer un statut « ressources humaines » plus noble que la supervision directe, par exemple, peut avoir sur ce dernier point des inconvénients évidents : surcharge mentale, épuisement...). Il vaudrait, donc, mieux parler de « symétrie de non-information ».

Comment pourrait se décliner, en interne, la T2A pour aboutir à une structuration qui aille dans le sens d'une performance améliorée ? Il semble que la granularité de l'outil, même avec 2 300 groupes de patients homogènes, n'est pas assez fine pour prendre en compte la diversité et l'incertitude qui pèsent sur les trajectoires de patients. En d'autres termes, le passage du résultat d'un CREA à ces dernières constitue un saut qui n'est en rien évident, mais cela signifie, alors, que cet outil ne peut pas constituer un outil de pilotage au niveau d'un pôle ou d'un service.

Les quelques efforts qui vont dans ce sens, se passent en fait des outils de la T2A, et se fondent sur des considérations pragmatiques, mais il convient de souligner combien ils sont difficiles, et que le découragement saisit souvent les acteurs locaux devant le défi d'avoir à coordonner des milliers de processus intriqués et aléatoires, et donc de gérer convenablement la « singularité à grande échelle » (Minvielle, 1996).

Ambiguïté du dialogue de gestion et création d'une fonction « organisateur »

Doit-on en conclure que la T2A est inopérante, ne donnant pas les moyens de s'attaquer à ce qui résiste profondément au cœur du métier, source de la performance, et qu'il faut (encore) passer à d'autres formes de régulation ? Ce n'est, sans doute, pas si simple.

Pour revenir au début de cet exposé et à la problématique de l'évolution des comportements, certes difficilement objectivable comme je l'ai dit, il semble malgré tout que la T2A soit bien à l'origine d'une proximité plus grande entre médecins et administratifs, surtout dans les établissements privés, où très souvent cette relation n'existait qu'à un état embryonnaire. De plus, une culture gestionnaire paraît réellement émerger au niveau des professionnels, comme le montrent les efforts d'un certain nombre de responsables de pôle ou présidents de commission médicale d'établissement (CME) que nous avons rencontrés. Il est, donc, à peu près certain que la T2A a entraîné une déformation des champs de jugement vers l'économique. Certes, d'après les constatations précédentes, elle aurait pour le moment plus de répercussions sur les représentations des acteurs que sur les modes concrets par lesquels ils s'organisent, mais rien n'interdit d'imaginer qu'à terme, ce basculement des perceptions n'embraye pas sur des effets concrets.

De son côté, le régulateur a pleinement conscience des inégalités de ressources entre les établissements et de l'existence de poches de sous-productivité dans le système hospitalier. Face à ces constats, il lui est difficile de renoncer à toute forme d'inscription gestionnaire dans les fonctionnements en cause, c'est-à-dire d'une prise effective de l'ensemble des dimensions de la performance par les opérateurs principaux des processus de soins. Compte tenu des immobilismes constatés avant l'avènement de la T2A, cette inscription ne peut se passer d'un incitatif économique, et d'une certaine façon les faits lui donnent raison au niveau des représentations des acteurs quant au souhaitable. Mais l'évaluation que nous avons effectuée, signifie essentiellement que cet ingrédient seul ne suffit pas si l'on veut que ce dialogue de gestion se traduise dans des organisations renouvelées.

Il nous semble que, parallèlement, l'organisation des processus de travail doit être prise en charge à la base, peut-être en créant des postes d'organiseurs ou en confiant cette mission à des médecins. La forme que cela pourrait prendre, demeure de toute façon peu claire pour l'instant. Cependant, des expériences qui ont été menées ici ou là, notamment dans le cadre de la MEAH (mission d'expertise et d'audit hospitalier), ont fait progresser les choses et montré et dessiné une méthode générale (MEAH *et al.*, 2008). Malgré la complexité de l'activité, et par l'intermédiaire d'un travail fin sur les processus mené en coopération avec les professionnels eux-mêmes, des dispositifs ont ainsi pu être modifiés, grâce à une coordination plus étroite, prenant mieux en compte les trajectoires de patients ; des durées de séjour ont diminué, ainsi que des délais de rendez-vous, des conditions de travail se sont améliorées, etc. De tels résultats ne sont reproductibles cependant qu'au niveau local, tant le système est complexe, et pose clairement les problèmes de capitalisation et de généralisation, mais il reste clair que si l'on veut progresser dans cette voie de l'inscription gestionnaire, le mouvement doit également partir de la base plutôt que de s'appuyer uniquement sur des principes de type *top-down*, comme ceux du financement des structures. Telle est, en tout cas, notre intime conviction, à l'issue de cette évaluation-intervention, dont on voit qu'elle ne fait pas qu'évaluer, mais qu'elle tente également, peut-être de façon aventureuse, d'imaginer les modalités d'une autre dynamique de changement. Parce qu'éprouver les champs de force de la situation, en compagnie des acteurs eux-mêmes, c'est aussi se donner les moyens d'y générer des impulsions, même légères, et d'expérimenter ainsi de nouveaux équilibres.

L'approche monographique

Hélène Genin (Euréval)

A lors que les séances précédentes de ce séminaire ont laissé une large place à l'approche quantitative, cet exposé s'arrêtera sur l'approche qualitative et sur un outil de collecte de données : l'étude de cas. Il s'appuiera sur l'exemple de l'évaluation des plans régionaux de santé publique. Une partie du questionnement de cette évaluation portait en effet sur l'appropriation des plans régionaux par les acteurs, de santé publique en particulier au niveau infra-régional.

S'approprier une politique publique

Politique et programme

Il est important de distinguer politique publique et programme public.

Cette dernière notion s'applique à un ensemble restreint d'objectifs censés faire sens pour les acteurs qui les ont conçus, pour ceux qui vont les mettre en œuvre ou ceux qui vont en bénéficier. Leurs objectifs ne font pas nécessairement débat. Par exemple, la réduction de la mortalité prématurée ne soulève aucune opposition. En revanche, lorsqu'un débat est soulevé autour de la légitimité du programme, les questions d'appropriation par les acteurs peuvent véritablement poser problème.

Un programme public se caractérise par des moyens humains, techniques, organisationnels et financiers dédiés, de même que par une temporalité limitée. L'approche retenue peut être au choix partenarial ou non. Dans le cas des plans régionaux de santé publique, la démarche relève pour une large part du partenariat. Ces plans sont pilotés par des groupements de santé publique dans lesquels les collectivités territoriales sont pleinement parties prenantes. Enfin, un programme est territorialisé. Les différentes collectivités qui composent le territoire vont alors participer chacune à leur niveau à la mise au point et à la réalisation du programme.

Une politique est un ensemble de programmes et s'inscrit dans un temps long. Il est donc plus difficile de se l'approprier. Une politique se caractérise par de nombreux changements. Le changement organisationnel n'est qu'une facette de ceux-ci, même s'il est souvent la

condition permettant d'atteindre des objectifs. Concernant les plans régionaux de santé, le cabinet Euréval est intervenu en Bretagne et en région Centre. Il intervient en ce moment en Bourgogne et en région Rhône-Alpes.

L'appropriation

Le terme d'appropriation renvoie à l'expression « faire sien », sur la base d'un système propre de valeurs et de représentations. Celles-ci diffèrent d'un acteur à l'autre, qu'il s'agisse d'un porteur de projet associatif qui mobilise des bénévoles ou d'un porteur de projet professionnalisé.

S'approprier implique en tout premier lieu de connaître. Dans le cas des plans régionaux de santé publique, il est plus simple de connaître leur contenu lorsque l'on a été associé à leur élaboration. Les stratégies de communication autour du plan vont cependant permettre aux acteurs de prendre connaissance de ce contenu. Souvent, en matière de communication, les intentions sont louables, mais les réalisations se limitent à un simple site Internet et quelques séminaires (encadré 1).

Ensuite, il faut que les acteurs comprennent le contenu et l'esprit du plan, qui rassemble généralement des objectifs de nature très différente. Certains peuvent être qualifiés d'objectifs de politique publique, en ce sens qu'ils ciblent des publics bien particuliers. D'autres, bien plus transversaux, relèvent en réalité plutôt de l'observation et du suivi des plans.

L'appropriation consiste aussi à utiliser les différents leviers d'action du programme. Les porteurs de projet peuvent s'appuyer sur les modalités du plan pour rechercher des subventions. Les animateurs territoriaux de santé peuvent s'en servir pour diffuser les objectifs régionaux et les réinjecter dans les initiatives locales. En effet, l'un des grands enjeux de l'élaboration de ces plans réside dans la mise en cohérence et le décloisonnement des interventions dans le secteur sanitaire, médico-social et social.

Le partage et la critique constituent le dernier stade de l'appropriation. La critique, même lorsque le contenu n'a pas été vraiment compris, s'apparente en effet à une forme d'appropriation, dans la mesure où elle met en jeu une distanciation par rapport aux pratiques propres. Elle permet d'alimenter la réflexion, d'apporter de nouvelles connaissances, de mettre en débat son interprétation propre et de contribuer à l'atteinte des objectifs plus généraux. Du point de vue de l'élu, certains objectifs nécessitent une communication plus douce que celle que propose le plan. Par exemple, les actions de prévention du suicide sont présentées en Bretagne comme un travail sur le lien social et le mieux-être.

L'appropriation des politiques publiques peut être feinte. Ce type d'attitudes se retrouve souvent chez les porteurs de projets, qui vont essayer de rentrer autant que possible dans les cadres. Ils vont bâtir des argumentaires autour des objectifs régionaux de santé publique. La vigilance doit donc être de mise et souvent il est nécessaire d'aller au-delà des simples discours, qui visent, de manière légitime, à collecter des fonds pour mettre en place ses interventions.

Il n'existe aucun indicateur d'appropriation. Celle-ci relève d'éléments de processus et l'approche quantitative est rarement la plus appropriée. Or cette notion détermine pour une large part les résultats d'une politique publique.

ENCADRÉ 1

L'appropriation des objectifs

Faire sien des objectifs – valeurs, représentations –, cela suppose de :

Connaître

Comprendre

Utiliser-décliner

Partager-critiquer

Alimenter (débat), contribuer (terrain)

L'évaluation des plans régionaux de santé publique

Les plans régionaux de santé s'appuient tout autant sur des appels à projet que sur des conventionnements. Les porteurs de projet, telles les associations, disposent d'une autonomie certaine. Cette autonomie se trouve tout de même limitée par l'existence du plan. En effet, c'est au titre de la mise en œuvre de la politique régionale et de sa cohérence qu'ont été choisis les porteurs de projets. Ainsi, s'instaure une tension permanente entre l'individualisation de la politique par les structures porteuses et la mise en cohérence de celle-ci. Il est important de voir dans quelle mesure ces acteurs sont à même de faire évoluer une politique.

Une évaluation des plans régionaux de santé publique a été entreprise à mi-parcours, conformément aux dispositions contenues dans la loi de santé publique de 2004, soit l'an dernier, puisqu'ils ont été lancés en 2006 pour quatre ans. Dans l'esprit du législateur, ces plans devaient marquer une avancée en matière de régionalisation de la politique de santé. Des enjeux de décloisonnement et de mise en cohérence des acteurs avaient été mis en évidence au moment de leur adoption. La complexité réside dans la multiplicité des champs et des intervenants sur le terrain, qui est une constante dans le domaine de la santé publique.

Les évaluations ont été cadrées au niveau national, sur la base d'un cahier des charges articulé autour de deux grandes questions. La première concernait la conception des plans, leur structuration et leur cohérence. Ceux que nous avons étudiés s'apparentent véritablement à des catalogues d'actions plutôt qu'à de véritables plans structurés. La deuxième question avait trait à la performance. En l'occurrence, l'évaluation porte plutôt sur les effets en termes de dynamique régionale plutôt que sur les impacts sur la population, qui fera l'objet d'une évaluation au terme du plan. Ainsi, c'est la problématique de l'appropriation qui se retrouve au cœur de l'étude au cours de cette première évaluation. Dans quelle mesure l'ensemble des acteurs (institutionnels, porteurs de projets, élus et représentants des usagers) ont-ils fait de ce plan un document de référence ? À ce propos, il faut noter que les plans associaient étroitement les représentants des usagers, au travers des conférences régionales de santé. Les partenariats, les modalités de travail et l'adéquation des actions financées avec les objectifs se retrouvent de fait également au cœur du dispositif d'évaluation.

**Évaluer l'appropriation :
l'apport d'une démarche qualitative par l'étude de cas*****Une enquête quantitative auprès des porteurs de projets***

Quels outils déployer pour mesurer l'appropriation ? Il serait vain d'opposer en la matière approche qualitative et approche quantitative. Ainsi, le dispositif retenu associe ces deux approches. Les évaluateurs du plan régional de santé publique de Bretagne ont essayé de

recueillir l'avis des porteurs de projets, qu'ils aient été retenus ou non. Une enquête en ligne très large a été lancée, afin de mesurer leur connaissance du plan régional, leur ressenti sur sa pertinence et l'utilisation qu'ils en faisaient. Étaient aussi évaluées par ces acteurs, dans le cadre de ce questionnaire, l'élaboration, la structuration et la lisibilité du plan. La deuxième partie était consacrée à la façon dont les porteurs avaient structuré leur projet, comment ils l'avaient financé, hors et à l'intérieur du plan. Le questionnaire s'attachait aussi sur les partenariats.

Au final, cette enquête a permis d'établir une photographie des perceptions, dessinant des tendances, notamment sur le niveau de connaissance du plan. L'approche ne visait pas à détecter d'éventuelles évolutions dans le temps, mais se concentrait sur l'aspect déclaratif. Des limites apparaissent dès lors de manière évidente. Ceux qui ont répondu à l'enquête sont bien évidemment les plus avertis et ce constat introduit un certain biais. Par ailleurs, deux questions ont permis de mettre à jour des incohérences. La première portait sur l'impact incitatif du plan régional en termes de partenariat. La très grande majorité des acteurs a répondu de manière positive. Cependant, une autre question visait à lister les partenariats noués par ces acteurs dans le cadre de ce plan. Or les réponses à celle-ci demeuraient peu nombreuses.

Ce manque de cohérence nous a démontré qu'il était impossible de se contenter d'une enquête quantitative pour évaluer l'appropriation. Toutefois, le recours à cette technique n'est pas inutile, dans la mesure où nous avons obtenu des précisions sur la manière dont les personnes avaient pris connaissance du plan et ce qu'elles en comprenaient. Cette première approche a donc été complétée par une seconde, à savoir une série d'études de cas. C'est cet aspect intégré dans nos propositions qui nous a d'ailleurs permis de faire la différence avec nos concurrents.

Complétée par des études de cas

L'étude de cas a pour but de vérifier comment sont déclinés concrètement les objectifs régionaux de santé sur les territoires. En effet, il est impossible de comprendre les processus de changement sans étude de terrain. L'intervention publique est structurée par des chaînes logiques. Ainsi, un faisceau de causes va produire un effet déterminé. L'enquête ne permet que faiblement de cerner ces mécanismes. Bien conçue, l'étude de cas permet de mettre à jour ces processus.

Selon les définitions des chercheurs américains, l'étude de cas se résume en l'observation d'un phénomène dans sa réalité et sa complexité (Yin). Concrètement, elle consiste à analyser une entité comme un tout dans son contexte. Nous avons retenu l'échelle du département comme la plus pertinente. Par ailleurs, au lieu d'observer l'intégralité du plan régional, nous nous sommes focalisés sur certains programmes particuliers : comme par exemple, la prévention du suicide dans les Côtes-d'Armor ; la lutte contre les addictions dans le département du Morbihan ; ou encore la prise en charge des maladies cardio-vasculaires dans le département du Finistère.

Chaque programme présente un statut différent. Certains avaient déjà fait l'objet de développement au niveau régional, tandis que d'autres n'étaient rien d'autre que la déclinaison au niveau régional de programmes nationaux. L'ensemble de ces variables a été intégré à l'étude.

Méthodologie

L'étude de cas procède d'une analyse holistique, s'appuyant sur un processus holographique. Nous espérons que l'analyse d'une partie nous permettra d'entrevoir l'ensemble. Il s'agit d'un objectif très ambitieux. Si le cas est mal sélectionné, il risque d'être difficile d'approfon-

dir l'analyse. L'étude de cas reste par essence empirique. Elle passe par l'analyse des données existantes. L'un des points forts de cette approche réside dans son aspect exhaustif, aussi bien au niveau quantitatif que qualitatif. Tous les points de vue des acteurs légitimes sont recueillis. Du côté des concepteurs, sont interrogés les référents thématiques du département concerné et les porteurs de projet, qu'ils dépendent des Comités départementaux d'éducation pour la santé (CODES) ou d'associations. Du côté des gestionnaires, sont interrogées l'Assurance maladie, l'Éducation nationale et les Directions départementales du travail, de l'emploi et de la formation professionnelle (DDTEFP).

Le but d'une telle étude réside avant tout dans le recoupement de l'ensemble des observations dans différents cadres. Elle a pour ambition de mettre en lumière les éléments structurants, permettant de comprendre les dynamiques à l'œuvre et les processus mis en place.

L'étude de cas peut porter sur une organisation, un groupe d'individus, voire un individu unique. Il peut aussi s'agir de partenariats, sachant qu'un tel dispositif n'est pas aisé à cerner. Les groupes et les réseaux peuvent aussi faire l'objet d'études de cas, de même que les projets. C'est ce dernier cas que nous avons retenu, puisque nous avons décidé d'analyser des programmes.

La méthode d'analyse se décline en plusieurs étapes.

La première consiste en la sélection des cas. Celle-ci ne relevait pas d'une démarche fermée, en chambre. Au contraire, les choix ont été réalisés avec l'instance qui commande l'évaluation, dans la mesure où il s'agit d'un groupe ressource, qui a connaissance à la fois des bonnes pratiques et des zones d'ombre dans les programmes. Parfois, cependant, les premières idées de cette instance ne se vérifient pas au cours de l'étude. Différentes variables sont croisées pour sélectionner les bons cas. Il est souvent préférable de s'adresser aux volontaires, qui veulent faire valoir certaines réalisations, faute de quoi les résistances sont trop importantes et les apports plus limités. De fait, ce sont ainsi souvent les éléments les meilleurs ou plutôt performants qui sont sélectionnés. Cette première étape s'avère cruciale, car elle conditionne le reste de l'étude.

La seconde correspond à la phase d'enquête proprement dite. L'enquête a pris la forme d'entretiens individuels ou de groupe, croisés avec un travail d'analyse documentaire. Sur cette base est rédigée une monographie. Robert K. Yin propose de structurer cette phase autour de deux étapes. Il estime que celui qui est en charge de la collecte de l'information n'est pas nécessairement le mieux placé pour rédiger la monographie. Il est vrai que cette première tâche risque de faire perdre une certaine distance au chercheur, qui parfois tend à entrer dans le jeu des acteurs. La monographie n'a pas pour ambition de produire un jugement, mais d'alimenter le jugement. Elle ne conclut pas, mais se contente d'apporter des constats.

Les apports des études de cas

Les études de cas ne sont pas déconnectées des questions de l'évaluation. Bien au contraire, elles permettent d'approcher une ou plusieurs problématiques. En croisant les critères, il est possible de réinjecter des éléments dans l'évaluation, ce qui constitue l'étape finale de cette démarche.

Cette méthode présente plusieurs points forts. Le premier réside dans le fait qu'elle permet de décrire les processus et de cerner les complexités qu'il serait impossible de mettre au jour sans cela. Par exemple, nous sommes allés interroger les porteurs des projets sur la manière dont le plan les incitait vraiment à nouer des partenariats et sur les freins qu'ils ont pu rencontrer dans leur ambition éventuelle. Parfois, les raisons n'avaient rien à voir avec le plan lui-même, mais tenaient à de simples conflits d'acteurs (encadré 2).

De plus, cette méthode aide à rendre compte de la réalité de la dynamique d'une organisation. Elle permet notamment de laisser s'exprimer des acteurs rarement interrogés. Dans le cas particulier du programme VIH et maladies infectieuses dans le Finistère, l'association Aides avait répondu au questionnaire, mais les personnes qui ont été rencontrées n'étaient pas les mêmes que celles qui avaient rédigé les réponses. Il s'agissait de bénévoles, qui disposaient d'une vision originale. Nous avons pu constater que le plan régional demeurait pour eux complètement abstrait, fort loin de leurs préoccupations. Seule la représentation régionale de l'association s'était véritablement investie sur ce sujet.

L'étude de cas alimente les rapports d'évaluation, met à jour des éléments concrets et permet d'identifier les logiques sous-jacentes de l'intervention qui ne sont pas explicitées *a priori*. Souvent, des hypothèses sont formulées avant même l'étude de cas, mais elles ne peuvent être véritablement étayées. Lorsque des éléments inattendus transparaissent, il est important de vérifier si les autres études de cas permettent de dégager une tendance similaire. Enfin, cette méthode permet de combiner trois éléments, à savoir la dimension managériale de l'évaluation, sa dimension cognitive et sa dimension pluraliste. La première met en évidence la place des acteurs dans les processus. La seconde est évidente. Toute étude participe en effet à l'accumulation des connaissances. La troisième trouve sa source dans le fait que l'étude de cas permet de recueillir l'ensemble des points de vue, en intégrant ceux qui bénéficient de l'intervention, en l'occurrence les porteurs des projets, subventionnés pour mettre en œuvre des actions à destination de publics spécifiques.

ENCADRÉ 2

Forces de l'outil

L'étude de cas :

Donne un aperçu des processus et des complexités qui sont impossibles à découvrir autrement

Rend compte de la réalité de la dynamique d'une organisation

Interroge des acteurs peu entendus (ex. : les bénévoles associatifs)

Fournit des exemples concrets, du réalisme

Permet d'identifier des logiques sous-jacentes de l'intervention

Leurs limites

Toutefois, comme tout outil, l'étude de cas présente des limites, assorties de difficultés pour sa mise en œuvre sur le plan pratique. En tout premier lieu, la généralisation peut s'avérer difficile. Il est difficile de s'appuyer sur des observations pour aboutir à une généralisation dans un champ où les acteurs se montrent particulièrement innovants. Le nombre restreint de cas étudiés rend cette étape d'autant plus complexe. De plus, il faut bien admettre que cet outil n'intervient qu'en complément d'études plus poussées et d'entretiens exploratoires.

Par ailleurs, l'étude de cas impose une tension entre une posture exploratoire, qui exclut dans la mesure du possible les *a priori*, et une posture confirmatoire. Souvent, l'étude de cas a vocation à confirmer les premiers résultats qui se dégagent de l'enquête. Cependant, il faut noter que certaines études de cas peuvent avoir une vocation purement exploratoire. Dans le cas des plans régionaux de santé publique, c'est une approche confirmatoire qui avait été retenue, tout en conservant un esprit aussi ouvert que possible.

Enfin, les biais inhérents à cette approche sont les mêmes que ceux généralement cités pour la conduite d'entretiens, notamment le biais d'empathie et la tentation qui pousse un chercheur à entendre ce qu'il a envie d'entendre. Seul le professionnalisme permet de les corriger. La rédaction de la monographie impose de conserver une certaine distance par rapport aux propos collectés.

Conclusion

En conclusion, je souligne que cet exposé se concentrait sur l'appropriation des objectifs des politiques publiques. Il aurait été possible de se pencher aussi sur l'appropriation des résultats de l'évaluation. La question renvoie à la définition des acteurs, la notion et la place des bénéficiaires. Dans le cas particulier des plans régionaux de santé publique, il fallait s'interroger sur la place des usagers dans l'appropriation et la conception des objectifs de santé publique. Se posaient en l'occurrence des questions autour des approches participatives, de leur finalité et de leurs résultats. Pour y répondre, une approche qualitative semblait indispensable.

SYNTHÈSE DES DÉBATS AVEC LA SALLE

Limites de l'intervention du chercheur

Question de la salle

Le terme même de recherche-intervention fait intervenir l'idée d'une participation active. Le chercheur n'est pas un simple observateur, mais il interfère avec les objets de la recherche. Quelle règle suit-il pour limiter son intervention de sorte à ce qu'elle ne se substitue pas aux démarches initiées par les objets observés ? Cette interrogation prend une importance particulière dans le cadre des projets de changement portés par un nombre restreint d'acteurs passifs ou réfractaires.

Réponse de M. Moisdon

Il est impossible de se dissocier entièrement de l'action des personnes observées. Dans de nombreux cas, le chercheur influe directement sur l'objet étudié. Dans d'autres, cette influence est limitée voire nulle. Dès lors, nous estimons que notre intervention n'a pas vraiment fonctionné.

Tout de même, certains dispositifs permettent de saisir l'impact du chercheur sur l'objet étudié. Ce dernier n'intervient pas pour présenter sa propre solution. Il s'inscrit dans un processus de changement préexistant et accompagne les acteurs dans leur plan. Il concourt à sa réalisation, qu'il l'estime ou non adaptée.

De plus, il existe plusieurs garde-fous méthodologiques. Le principal réside dans le nombre de chercheurs impliqués dans une même intervention. Jamais un chercheur n'intervient seul. Dans le cas de la tarification à l'activité, ce sont quatre chercheurs qui travaillaient ensemble. Cela permet de confronter les points de vue, car généralement les lectures d'un même événement diffèrent d'une personne à l'autre. Ainsi, un contrôle mutuel s'instaure. L'implication se gère de manière collective. Des réunions sont organisées à intervalle régulier pour évoquer les recherches menées par les autres. Les chercheurs s'interrogent entre eux pour recadrer leur activité et émettre des doutes sur les actions entreprises. Une gestion fine se met ainsi en place.

Pour autant, il n'est pas simple de ne pas s'impliquer au beau milieu du travail des autres acteurs. Cependant, ce n'est pas l'intervention en elle-même qui s'avère intéressante, mais uniquement ce qu'elle révèle de l'organisation. Les chercheurs ressentent en effet directement la force réactive.

Dans le cas d'un groupe pétrolier qui avait sollicité notre assistance, le modèle théorique choisi s'avérait particulièrement sophistiqué et supposait une organisation complètement intégrée dans laquelle les acteurs dialoguent en permanence les uns avec les autres. Or le groupe présentait une organisation en râtelier classique, qui freinait considérablement la mise en place du modèle cible. De fait, ces obstacles révélaient la nature réelle de l'organisation, au-delà des discours officiels très maîtrisés. Nous avons ainsi mis en évidence un cloisonnement extraordinaire et une activité bureaucratique prégnante. Nous avons exposé aux responsables les conclusions de notre étude. Ils les ont acceptées. Pour autant, ils n'ont jamais su corriger l'organisation pour mettre en place le modèle.

Effets des interventions dans la durée

Question de la salle

Je me demandais si, dans le cadre de vos actions de recherche-intervention, il vous arrivait de pratiquer des évaluations à terme, un an ou deux ans après la fin de votre intervention. Comment vous assurez-vous de l'acculturation et de la mise en place effective ? Comment pouvez-vous être certains qu'un élément nouveau résiste à l'épreuve du temps ?

Réponse de M. Moisdon

Cette question est assez cruelle, car, de manière générale, nous ne sommes pas vraiment bons dans ce domaine. Parfois, nous ne prenons pas le temps d'observer ce qui reste à terme d'une intervention.

Nous l'avons cependant fait sur le dossier des hôpitaux. Nous avons, dans une large mesure, participé à la construction des PMSI. À l'origine, étant mathématicien de formation, je travaillais sur l'élaboration de modèles pour ces structures. Ces créations étaient accueillies avec un véritable enthousiasme et, au final, les résultats étaient quasi nuls. Nous nous sommes alors aperçus qu'il manquait un aspect incitatif majeur. Lorsque Jean de Kervasdoué a importé l'idée du PMSI des États-Unis pour l'adapter à la France, nous avons accouru. Par la suite, nous avons vu les problèmes se multiplier.

Dans le cas de Renault, nous avons travaillé 12 ans sur le processus de conception des voitures. Nos apports sont en l'occurrence importants. Les dirigeants de Renault étaient persuadés de l'existence d'un cloisonnement excessif au niveau des techniciens. Nous leur avons démontré au contraire l'existence d'un véritable dialogue à ce niveau. En réalité, c'étaient les savoirs techniques qui posaient problème.

D'autres interventions s'avèrent beaucoup plus problématiques. Ainsi, je ne suis pas certain des résultats obtenus au travers de notre intervention auprès des Caisses d'allocations familiales. Nous avons travaillé sur le contrôle de gestion sans grand succès. Nous sommes prêts à reconnaître que, sur certaines interventions, il aurait été nécessaire de procéder à une évaluation à terme, que nous n'avons pas menée.

Légitimité de l'intervention

Question de la salle

Il est évident que l'on n'observe pas sans intervenir. Cependant, se pose la question de la légitimité de l'intervention. À quel titre et dans quelle mesure un intervenant extérieur peut-il intervenir ? Dans le domaine de l'action sociale, le recours à des monographies est très fréquent. Je pense qu'il convient alors de s'intéresser à la fois aux domaines où des réalisations importantes ont été identifiées mais aussi et surtout à ceux où des échecs retentissants ont été enregistrés. Même si le travail de terrain est plus complexe dans le deuxième cas, il permet de mettre en évidence des éléments très intéressants.

Réponse de Mme Genin

Nous n'avons pas véritablement poussé l'analyse sur les postures exploratoires. Lorsque les processus et les hypothèses demeurent mal assurés, il est très intéressant d'utiliser

une monographie. Dans le cadre d'évaluations sur des durées très limitées, nous allons à l'essentiel, à savoir les exemples les plus efficaces. Il est plus simple d'obtenir des résultats ainsi. Cependant, si nous disposions d'un temps plus long pour mener nos études, nous nous pencherions sur des échecs patents pour enrichir nos conclusions. Dans ce cadre, la recherche-intervention représente une méthode particulièrement opportune.

Réponse de M. Moisdon

Il faut bien souligner le fait que nous ne démarchons jamais les organisations. Nous ne faisons que répondre à leurs demandes. Nous disposons avant toute chose d'une légitimité technique. La gestion s'appuie sur des indicateurs, une comptabilité analytique et générale et des modèles. Nous pouvons faire valoir des compétences fortes en la matière. Nous nous présentons dès lors comme des techniciens qui entreprennent une interaction instrumentée.

Face à un projet, nous proposons des indicateurs de suivi. Nous avons accumulé une masse de connaissances importante, au travers notamment de lectures savantes. Cependant, nous avons des lacunes réelles dans certains domaines comme le marketing. Au-delà, nous ne nous contentons pas d'établir des tableaux de bord. Nous observons leurs effets, indésirables notamment. Les entreprises nous connaissent pour notre capacité à analyser en profondeur les grilles de gestion.

Intervenir, c'est toujours intervenir sur la vie des autres. Sur la base de ce constat, il faut prendre position. Si nous acceptons la mission, nous devons en peser toutes les conséquences.

Différence entre instrument et politique publique

Question de la salle

Je n'ai pas l'impression qu'il soit possible de comparer la tarification à l'activité à une politique publique. En effet, une politique publique vise un effet sur les bénéficiaires, à savoir les usagers de l'hôpital. La tarification à l'activité s'apparente davantage pour moi à une stratégie.

Réponse de M. Moisdon

Nous ne nous sommes pas demandés si la tarification à l'activité relevait véritablement d'une politique publique. Pour nous, il s'agit tout simplement d'un objet fascinant qui rentre pleinement dans le cadre de nos investigations. Peut-être s'agit-il d'un monstre plutôt que d'une politique publique. En tout état de cause, je me demande si la tarification à l'activité rentre dans le cadre d'une politique publique. Tout un discours vise en effet à la rattacher à une politique publique liée au *public management*.

Nous considérons les instruments tels qu'ils sont, en ne négligeant pas le fait qu'ils interagissent avec d'autres éléments. Dans le cas de la tarification à l'activité, nous pourrions citer la nouvelle gouvernance. Parfois, il est difficile de distinguer les effets de chacun. Dans le domaine de la contractualisation interne, en revanche, il semblerait que la nouvelle gouvernance influe plus directement que la tarification à l'activité.

Pour nous, la notion de bénéficiaire constitue un élément abstrait. Nous nous intéressons avant tout aux médiations gestionnaires. Nous nous sentons totalement incompétents

pour évaluer l'impact d'un tel outil sur les bénéficiaires. La méthode de la recherche-intervention n'aurait aucun sens dans ce cadre, pour lequel il est nécessaire de mettre en œuvre des méthodes plus classiques.

Réponse de Mme Genin

Selon moi, la tarification à l'activité ne relève pas d'une politique publique, ni même d'un programme. Il s'agit d'un outil ou d'un instrument imposé au service d'une politique plus globale.

Comment articuler évaluation d'une politique publique et évaluation d'une organisation ? Les approches et les objets diffèrent, selon moi. Une politique publique produit des changements sur les bénéficiaires mais aussi sur les opérateurs. Son évaluation est génératrice de changement et remet du sens dans l'intervention. Les acteurs sont en effet amenés à s'interroger sur leurs postures. Dans ce cadre, apparaît un lien direct avec l'organisation.

Exhaustivité des données dans le cadre de l'étude de cas

Question de la salle

Je suis très intéressée par cette méthode qui consiste à concentrer son regard sur quelques objets pour traduire des tendances plus larges. Cependant, je m'interroge sur la question de l'exhaustivité des données recueillies dans le cadre de l'étude de cas.

Réponse de Mme Genin

Concernant l'exhaustivité des données, je tiens à préciser que le champ de notre analyse ne portait pas sur les résultats épidémiologiques auprès des publics. L'intérêt de l'approche quantitative s'avérait donc particulièrement limité dans ce contexte. Nous avons cependant évalué le nombre de partenariats et leur durée. Dans d'autres situations, notamment les études de cas sur les entreprises bénéficiaires de financements publics, une approche quantitative plus large s'impose pour mieux cerner les phénomènes et les résultats. Tout dépend en réalité de l'objet étudié. Nous visons l'exhaustivité pour comprendre tous les aspects qui influent sur un domaine particulier. Cependant, pouvons-nous réellement prétendre à l'exhaustivité ?

Il est envisageable de commencer par les études de cas, afin d'élaborer ensuite un questionnaire. Cependant, cela aurait signifié que les études de cas avaient une vocation essentiellement exploratoire. Ainsi, nous aurions cherché uniquement à confirmer des hypothèses qui seraient apparues dans le cadre des études de cas. Une telle approche aurait été en l'occurrence certainement moins riche.

Il est inutile de lancer une enquête lorsque l'on connaît mal le public à qui elle s'adresse. Grâce à nos entretiens avec les concepteurs des plans et les référents thématiques des programmes, nous disposons d'une bonne connaissance des porteurs de projets.

BIBLIOGRAPHIE

- Bazzoli G., Dynan L., Burns L., Yap C.**, 2004, « Two decades of organizational change in Health care : What have we learned ? », *Medical Care Research and Review*, vol. 61, n° 3, p. 247-331.
- Berry M.**, 1983, *Une technologie invisible*, Paris, École polytechnique.
- Desremeaux A.**, 2005, *Théorie des organisations*, Paris, Éditions EMS.
- Engel F., Kletz F., Moisdon J.-C., Tonneau D.**, 2001, *La démarche gestionnaire à l'hôpital, t. II, le PMSI*, Paris, Seli Arslan.
- Glouberman S., Mintzberg H.**, 2001, « Managing the Care of Health and the Cure of Disease », *Health Care Manage Review*, p. 56-86.
- Grimand A., de Vaujany F.-X.**, 2006, « Repères pour un management de l'appropriation des outils de gestion », in **Lièvre P., Lecoutre M., Traoré M.** (dir.), *Management de projet*, Éditions Hermès.
- Hatchuel A.**, 2000, « Research, intervention, and the production in knowledge », in *Cow up a tree* (collectif), Éditions INRA.
- Meah et centre des gestion scientifique de l'École des Mines de Paris**, 2008, *Organisation hospitalière, approche et expériences*, Berger Levrault.
- Minvielle E.**, 1996, *Gérer la singularité à grande échelle. Comment maîtriser les trajectoires de patients à l'hôpital ?*, thèse de doctorat de l'École polytechnique, Paris.
- Mintzberg H.**, 1982, *Structure et dynamique des organisations*, Paris, Éditions d'Organisation.
- Moisdon J.-C.**, 1997, *Du mode d'existence des outils de gestion*, Paris, Seli Arslan.
- Moisdon J.-C.**, 2006, *La recherche intervention : de la prescription à l'énigme*, Intervention et savoirs, Conservatoire National des Arts et Métiers, Paris
- Moisdon J.-C.**, 2008, « Gouvernance clinique et organisation des processus de soins : un chaînon manquant ? », *Processus et Organisation des Soins*, vol. 39, n° 3, p. 175-181.
- Lenay O., Moisdon J.-C.**, 2003, « Du système d'information médicalisée à la tarification à l'activité : trajectoire d'un instrument de gestion du système hospitalier », *Revue française de gestion*, n° 146, p. 131-141.
- Reynaud J.-D.**, 1998, *Les règles du jeu*, Paris, Armand Colin.
- Yin R. K.**, 1984, *Case study research : Design and methods*. Newbury Park, CA, Sage.

ASSOCIER LES USAGERS ET LES CITOYENS

La quatrième séance du séminaire sur les méthodes d'évaluation de politiques publiques s'est intéressée à la prise en compte du point de vue des bénéficiaires dans l'évaluation des politiques publiques.

– Deux économistes, Francine Fenet et Audrey Daniel, du Pôle politiques publiques et prospective de la ville de Nantes, ont dans un premier temps fait part des expériences d'évaluation conduites à Nantes en mettant l'accent sur la façon dont les besoins et les attentes des populations y ont été intégrés.

– Dans un deuxième temps, Sylvain Ortis, chargé de mission en évaluation des politiques publiques au conseil régional Rhône-Alpes, a présenté les pratiques de la région concernant l'association des bénéficiaires et des citoyens au processus d'évaluation.

Ces deux interventions ont été suivies d'un débat avec la salle.

L'importance de la prise en compte des besoins et attentes de la population dans les évaluations

Francine Fenet et Audrey Daniel (ville de Nantes)

Contexte, orientations et partis pris depuis 2004

L'évaluation s'est retrouvée au cœur du projet managérial « Nantes s'engage », visant à faire évoluer le pilotage et les pratiques de l'action publique. Plutôt qu'une évolution des pratiques professionnelles, ce projet managérial avait pour but de prendre en compte de manière plus précise des besoins et attentes des Nantais sur l'ensemble du champ de compétence de la Ville (18 politiques publiques depuis la petite enfance, jusqu'aux personnes âgées en passant par la politique culturelle). Toutes les politiques y sont décrites, en reprenant le contexte, les finalités, les plans d'actions et orientations stratégiques. Il s'agit donc d'un cadre précieux pour référencer les démarches d'évaluation des politiques publiques.

Dans sa lettre de mission au directeur général des services, le maire de Nantes demandait plus de visibilité sur les résultats et effets des actions menées. Ces démarches devaient s'inscrire dans le cadre d'une nouvelle gouvernance locale, à même d'éclairer les arbitrages politiques dans le but de corriger et d'améliorer les actions, les services et les prestations. Ainsi, un accent tout particulier a été mis sur les préconisations et leurs mises en œuvre.

L'évaluation au sein de la municipalité : une démarche volontariste, interne, pilotée et animée par la Mission « évaluation »

La charge de l'évaluation a été confiée à un groupe constitué d'un service dédié (la Mission évaluation) et de personnes ressources dans différentes directions municipales. Le parti pris était de développer et d'accompagner des démarches d'évaluation à visée démocratique, telles que les qualifie l'Association des départements de France en 2008. Ces travaux s'inscrivent d'ailleurs dans le droit fil de ceux menés autour de la performance et de la qualité pour la démocratie locale.

Ainsi, c'est une démarche volontariste qui s'est mise en place en interne, tant la Mission évaluation demeure présente dans l'accompagnement du processus. L'animation du dispositif demeure en effet à sa charge. Le recours à des prestataires extérieurs est centré sur la collecte de données et la réalisation d'enquêtes. Chaque année est défini un programme de tra-

vail présenté en Bureau municipal. Ce sont les domaines stratégiques à forts enjeux en termes de besoins et d'attentes qui sont privilégiés ; ou encore des services ou des politiques autour desquels des faiblesses ont été repérées. Par exemple, la politique à l'égard des personnes âgées se concentrait jusqu'à récemment quasi exclusivement sur le placement en établissement, alors que les Nantais réclamaient des actions fortes en faveur du maintien des personnes âgées à domicile. Il a fallu alors redéfinir la politique de la Ville dans ce sens. Un questionnement auprès des personnes âgées a mis notamment en évidence des problématiques d'isolement et d'adaptabilité des logements. Sur ces points, la Ville demeure en première ligne et particulièrement pertinente dans ses interventions.

La Mission évaluation demeure proche des directions municipales dans le processus d'accompagnement. Un lien étroit est maintenu en parallèle avec les élus, comme cela est souvent le cas au niveau local. Le processus est par essence pluraliste et essaie de croiser dans la mesure du possible les points de vue des professionnels et des usagers. Le Comité opérationnel de l'évaluation auquel est adossée la Mission permet de conserver une certaine distance par rapport aux évaluations menées, aux méthodes et à la formulation de préconisations, malgré le positionnement de la Mission au cœur même des services de la Ville. Le processus s'avère au final assez fructueux. Les acteurs s'approprient désormais de manière collective la démarche d'évaluation des politiques publiques.

La méthodologie retenue

Du point de vue méthodologique, la démarche d'évaluation cherche à capitaliser de manière systématique les connaissances existantes, produites par les grandes enquêtes de l'Institut national de la statistique et des études économiques (INSEE) au niveau national, mais aussi par la région et le département. La ville de Nantes observe ainsi les pratiques sur le reste du territoire français, avant d'engager ses propres investigations. Au-delà de ce travail de synthèse documentaire, une personnalité experte du sujet, généralement un chercheur, est associée aux travaux afin de resituer les politiques publiques de la Ville dans un contexte plus global.

Ensuite, les services essaient dans la mesure du possible de recenser l'ensemble des effets induits par les politiques publiques. L'enquête s'appuie en général sur différents volets complémentaires qui permettent de saisir la réalité dans sa complexité. Par ailleurs, nos services ne sont pas attachés à une technique d'évaluation particulière. Toutes les méthodes (enquêtes quantitatives, qualitatives, analyse cartographique, diagnostics et modélisations) sont explorées et il est fait appel à diverses disciplines scientifiques qui peuvent apporter un éclairage sur les politiques évaluées, en fonction des questions qui se posent. Des disciplines voisines de l'évaluation peuvent être appelées en renfort, à l'image des outils de gestion.

Parallèlement, nous réinterrogeons les publics cibles sur leurs besoins et leurs attentes. Ces données étant rarement disponibles, il nous faut les produire nous-mêmes. De plus, des évolutions peuvent intervenir assez rapidement. Il faut donc procéder de manière régulière à des enquêtes, qui fournissent une image plus vivante de la société.

Des comparaisons sont en outre lancées pour vérifier si les résultats obtenus diffèrent de ceux observés par ailleurs, au niveau national ou dans d'autres collectivités. Nous nous appuyons par conséquent sur des questionnaires existants pour établir nos propres questionnaires. Dans le cas de la politique culturelle, nous avons pris pour base de travail l'enquête sur les pratiques culturelles des Français de l'INSEE et celles du Centre de recherche pour l'étude et l'observation des conditions de vie (CRÉDOC) sur les bibliothèques locales. Ces questionnaires sont travaillés de manière collective. Les directions et les partenaires sont donc associés directement, de sorte à ce que le processus et les conclusions deviennent plus lisibles et que les résultats soient moins remis en cause au moment de la restitution. Cela

permet d'obtenir une meilleure appropriation des résultats.

Au moment du travail sur les préconisations, nous nous nourrissons des expériences menées par d'autres collectivités, pour les mettre en perspective. Avant de proposer de reprendre une initiative mise en place par une autre collectivité qui nous semble pertinente, nous essayons de mesurer les résultats obtenus par celle-ci. Il s'agit de tirer le meilleur parti des expériences menées par d'autres.

Enfin, comme nous voulons assurer une large diffusion des résultats de l'évaluation auprès de l'ensemble des parties prenantes, il faut croiser et confronter les points de vue des citoyens, des professionnels, des partenaires et des élus. Même si les intérêts sont parfois contradictoires, c'est ainsi que se dessinent les préconisations les plus pertinentes.

Illustrations

La politique en faveur de la petite enfance

Avec le conseil général et la caisse d'allocations familiales (CAF) de Loire-Atlantique, une évaluation a été menée entre mars 2006 et octobre 2007 sur l'accueil à domicile par les assistantes maternelles agréées indépendantes. La politique publique municipale en la matière avait pour objectif d'améliorer la capacité et la qualité des modes d'accueil individuels, pour l'adapter aux besoins des familles, au moment même où nous constatons une tension très forte au niveau des structures collectives municipales. Beaucoup de demandes restaient en effet insatisfaites et les familles étaient franchement frustrées. Par ailleurs, certaines assistantes maternelles agréées indépendantes connaissaient une certaine inactivité. Cela nous donnait l'impression que leur offre d'accueil n'était pas bien adaptée aux besoins des familles.

Nous avons donc cherché à objectiver la situation à travers un diagnostic partagé. Nous nous sommes entendus autour de trois grandes questions. La première portait sur la mesure quantitative de l'adéquation entre l'offre « petite enfance » et les besoins des familles. La seconde avait trait à une appréciation qualitative de l'offre de service proposée par les assistantes maternelles. Enfin, la dernière question concernait la place du Relais assistantes maternelles (RAM), géré par la Ville, dans le cadre de la mise en relation entre assistantes maternelles et familles demandeuses. Les partenaires voulaient ainsi interroger les leviers d'action qui permettent de favoriser l'adéquation entre l'offre et la demande.

Le pilotage de l'évaluation

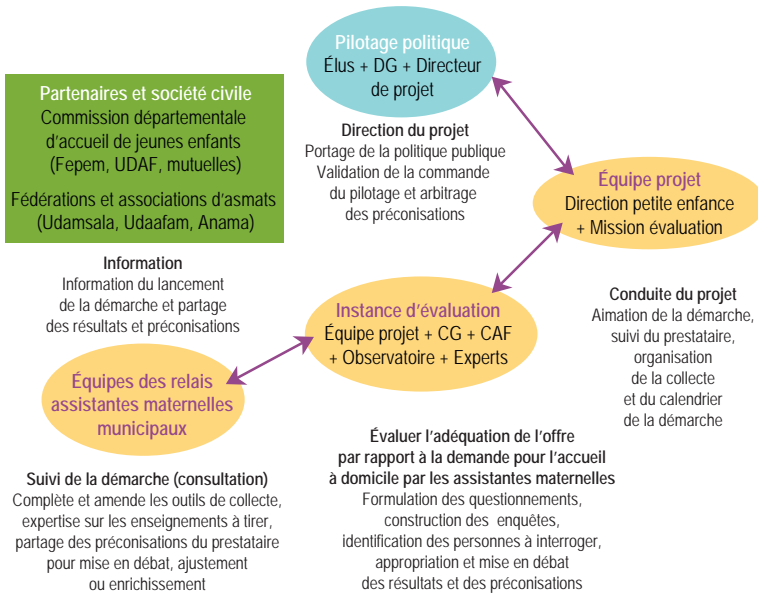
L'instance d'évaluation mise en place a travaillé sur les outils de collecte des données, l'appropriation des résultats et les préconisations. Dans le cadre de ce projet, cette instance a associé étroitement les différents partenaires impliqués dans cette politique, en l'occurrence la responsable de la protection maternelle et infantile du Conseil général et le Directeur de l'action sociale de la CAF. Catherine Sellenet, chercheuse au CNRS, qui a beaucoup travaillé sur les assistantes maternelles, a été choisie comme experte. Ce petit groupe a travaillé sur les questions évaluatives, sélectionné un partenaire chargé de l'enquête de terrain, suivi les résultats, construit les outils de collecte et formulé les préconisations (schéma 4).

En parallèle, la démarche a été partagée avec les équipes des Relais assistantes maternelles municipaux notamment. Elles étaient ainsi mises au courant de l'avancement des travaux, dans une optique de transparence. Les autres partenaires ont également été informés de l'existence de cette démarche, à travers la Commission départementale de l'accueil du jeune enfant, qui regroupe la Fédération des particuliers employeurs (FEPÉM), l'Union départementale des associations familiales (UDAF) et les Mutuelles de Loire-Atlantique, qui gèrent

des crèches associatives. Enfin, les fédérations d'assistantes maternelles ont elles aussi été associées, pour compléter la dynamique. Un pilotage politique a également été mis en place avec les élus référents, tout en maintenant un lien fort avec la direction en charge des Relais assistantes maternelles.

SCHÉMA 4

Le dispositif de pilotage et de suivi de l'évaluation accueil à domicile par les assistantes maternelles



L'organisation de la collecte des données

Après mise en concurrence, TMO-Régions/GERS a été sélectionné sur la base d'une proposition déclinée en plusieurs volets. Différents modules devaient permettre de répondre à un questionnaire complexe recouvrant plusieurs problématiques différentes. Au départ, il s'agissait de recenser les places disponibles et le nombre d'enfants pour lesquels les familles avaient besoin d'un accueil pérenne. Un diagnostic a ainsi été établi quartier par quartier. Cette première phase a été complétée par un sondage téléphonique auprès de 400 familles comptant au moins un enfant de moins de trois ans, sur la base des fichiers de la CAF de Loire-Atlantique. Certaines d'entre elles avaient recours régulièrement à une assistante maternelle, tandis que d'autres avaient abandonné cette solution. D'autres encore n'y avaient jamais eu recours. De la sorte, nous étions en mesure de réobjectiver les pratiques globales et les besoins, ainsi que la perception autour de ce mode de garde.

Ces éléments ont été complétés par une enquête postale exhaustive auprès des 1 256 assistantes maternelles indépendantes agréées sur le territoire nantais, avec un taux de retour supérieur à 60 %, ce qui était nettement au-dessus de nos attentes. Nous voulions mieux cerner les pratiques et la place du Relais assistantes maternelles.

En outre, 15 entretiens individuels ont été menés avec des parents qui ne parvenaient

pas à trouver un mode de garde. Ces familles ont été identifiées via les RAM. Nous voulions en effet étendre nos investigations au-delà du seul cercle des usagers pour y inclure des personnes exclues de ce service, mieux comprendre les difficultés d'accès et déterminer si celles-ci relèvent de problèmes d'information, d'accompagnement ou d'ordre financier.

Enfin, quatre tables rondes ont été organisées :

- deux autour d'assistantes maternelles de plus et de moins de 35 ans, car il était apparu dans nos premières données un positionnement professionnel différent entre les plus jeunes et les plus âgées ;
- deux avec les responsables de Relais assistantes maternelles.

Les résultats

Il est ressorti de ces études que six familles sur dix privilégiaient l'accueil collectif, ce qui explique la tension constante autour de l'offre de la Ville. Les entretiens individuels ont permis de montrer que les personnes qui venaient d'arriver dans la Ville se trouvaient dans une situation plus difficile que les autres, de même que les demandeurs d'emploi.

Cependant, une famille sur quatre privilégie dès l'origine l'accueil individuel. Cette solution souffre tout de même d'un déficit d'image, expliquant la sous-activité récurrente des assistantes maternelles, notamment des plus jeunes ou implantées dans un quartier où l'habitat social domine. Ainsi, nous avons déterminé que 200 places demeuraient inutilisées. En 2008, ce chiffre semblait s'être fortement réduit, avant de repartir à la hausse dernièrement. Cette dernière tendance s'explique par la crise et les efforts du Conseil général pour développer l'offre. Ces évolutions rapides nous contraignent à réactualiser le diagnostic de manière régulière.

Nous avons noté par ailleurs que l'offre d'accueil des assistantes maternelles indépendantes s'avérait stable. Les plus jeunes s'inscrivent véritablement dans une optique de professionnalisme, ce qui nous incite à investir sur cette solution, d'autant que le niveau de satisfaction des usagers s'avère très élevé. La souplesse qu'offre cette solution est en effet grandement appréciée par les familles qui ont des contraintes professionnelles lourdes.

Nous avons conclu sur le fait que l'offre individuelle était à l'heure actuelle insuffisamment mobilisée. Le Relais assistantes maternelles est jugé comme efficace aussi bien par les familles que par les assistantes maternelles. Cependant, il ressortait des études que ses missions pouvaient être élargies.

Les partenaires ont travaillé sur un ensemble de préconisations avec, pour but premier, de restaurer la confiance et la crédibilité, au moyen d'une communication cohérente. Il a ensuite été proposé de renforcer la fonction d'information et de conseil des RAM et de développer un accompagnement particulier en direction de certaines familles, comme les nouveaux arrivants ou les familles présentant des situations professionnelles instables. Il était aussi recommandé de développer le conseil à la fonction employeur, en lien avec la FEPEM. En coordination avec le Conseil général et les fédérations d'assistantes maternelles, nous avons proposé l'organisation d'ateliers autour de questions relatives aux enfants (sommeil des enfants, secourisme...). Nous avons préconisé aussi la mise en place de passerelles plus nombreuses avec le collectif, au travers notamment d'ateliers et d'activités partagées dans les structures collectives impliquant les assistantes maternelles. Cette action visait en priorité les quartiers où domine l'habitat social. Par ailleurs, nous avons proposé d'établir un lien avec l'accueil d'urgence qui permet de répondre aux problématiques de reprise d'emploi ou de maladie. Enfin, nous avons recommandé le développement d'actions de soutien à l'offre, en visant plus particulièrement certains quartiers et en repérant les assistantes maternelles qui se positionnent sur des créneaux plus spécifiques (horaires décalés, handicap...).

La politique de lecture

Une autre évaluation intéressante a été menée sur la politique de lecture de la Ville, avec pour objectif d'optimiser le service rendu, en prenant en compte les nouveaux besoins. Il s'agissait de faire de la lecture un bien partagé et d'offrir à tous les publics un meilleur accès aux différents lieux de lecture et d'expression. Dans son programme, en 2001, le maire appelait à renforcer la politique de proximité dans les quartiers et à promouvoir les équipements culturels comme levier de médiation pour tous les publics. Parallèlement, nous constatons un basculement dans la société internet et la transformation des bibliothèques en médiathèques, qui nécessitait un ajustement de la politique première.

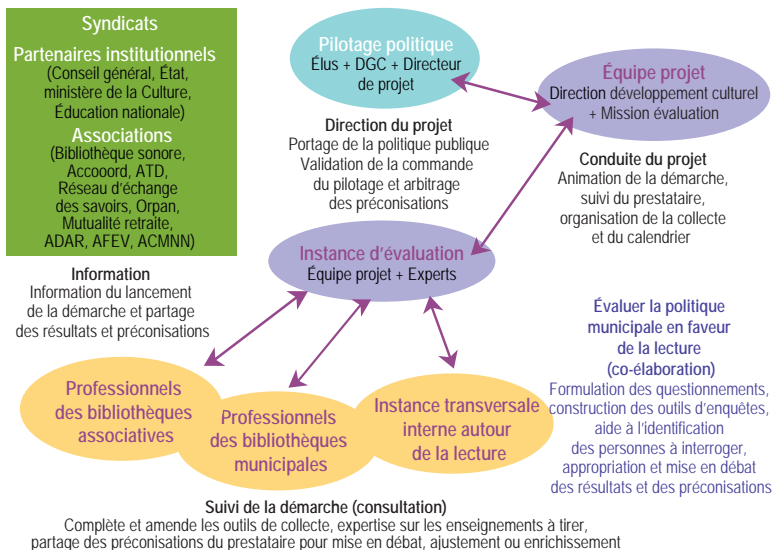
En tout premier lieu, nous avons cherché à identifier les pratiques et les usages des Nantais vis-à-vis des lieux de lecture et des actions, qu'elles soient municipales ou associatives, et à recenser les besoins et les attentes en matière d'accès à la lecture et au livre (besoins couverts et non couverts par l'offre de lecture existante). Plus encore, nous souhaitons identifier les complémentarités dans le maillage de l'offre municipale et associative.

Le pilotage de l'évaluation

C'est un dispositif de pilotage assez complexe qui a été mis en place (schéma 5). L'évaluation a cette fois été pilotée en interne en associant la directrice des Bibliothèques municipales, l'animatrice de la politique lecture et la Mission évaluation. Nous avons en outre eu recours à des instances d'animation de la politique lecture pour informer et consulter sur cette démarche. Nous avons enfin multiplié les échanges avec les professionnels des bibliothèques municipales et des structures associatives, mais aussi avec les autres acteurs de la Ville potentiellement impliqués, comme la Direction des personnes âgées et la Direction de la jeunesse, notamment sur les questionnaires. Un volet d'information a été prévu en direction de tous les partenaires associatifs et institutionnels.

SCHÉMA 5

Le dispositif et de suivi pour l'évaluation « lecture »



Une évaluation en quatre phases

TMO-Régions a été choisi pour produire les éléments de connaissance, après mise en concurrence.

Au cours d'une première phase (décembre 2008-avril 2009), nous avons analysé les pratiques et l'adéquation entre l'offre et la demande, par le biais d'une enquête téléphonique auprès de 1 100 Nantais âgés de 16 ans et plus. Cet échantillon se trouvait être suffisamment large pour établir une cartographie sociale des pratiques de lecture et de fréquentation des équipements. A été lancée, par ailleurs, une seconde enquête auto administrée auprès des usagers des lieux de lecture (bibliothèques municipales ou associatives), afin de dresser une typologie des usages et des pratiques en fonction des différents groupes sociaux qui fréquentent les bibliothèques, mais aussi d'offrir une vision plus large sur l'offre. Une table ronde rassemblant douze usagers de la bibliothèque municipale a été, par ailleurs, organisée pour évoquer les pistes d'amélioration. Nous avons en outre procédé à un important travail d'analyse de nos fichiers de gestion.

La deuxième phase (mai à septembre 2009) était centrée sur l'évaluation des besoins, attentes et actions engagées auprès des publics spécifiques. Une douzaine d'adolescents ont ainsi été envoyés deux par deux dans chacune des bibliothèques de la ville pour faire part ensuite, lors d'une table ronde, de leur ressenti vis-à-vis de l'offre. S'agissant des personnes âgées, ce sont les professionnels qui ont engagé des actions en direction de ces usagers qui ont été interrogés. Nous avons préféré nous tourner pour cette partie vers des professionnels plutôt que vers des usagers parce que ce champ reste mal couvert. C'est le même choix qui a été retenu pour les publics éloignés de la lecture, ou repérés ainsi par l'enquête téléphonique. Une table ronde a été organisée autour d'une dizaine de professionnels. Parallèlement, une quinzaine d'entretiens individuels ont été menés avec des personnes qui s'étaient inscrites gratuitement à la bibliothèque auprès du Centre communal d'action sociale (CCAS) à l'occasion de l'initiative « Lire en fête ». Nous avons pu ainsi vérifier si ces personnes avaient pérennisé leur pratique et repérer d'éventuels problèmes. Une dernière table ronde a été organisée autour de la question de l'accès des enfants à la lecture en dehors de l'école, réunissant des acteurs professionnels ou associatifs intervenant sur les temps périscolaire et extrascolaire.

La troisième phase consistait en l'élaboration des préconisations (octobre 2009). Celle-ci a débuté par un séminaire d'une journée d'appropriation de l'ensemble des résultats collectés. Le prestataire a organisé parallèlement une journée de travail avec 15 professionnels pour partager ces résultats et mettre en débat les préconisations du comité de pilotage, les ajuster, les enrichir et les hiérarchiser.

D'octobre 2007 à juillet 2008 avait été, en outre, menée une évaluation des bibliothèques-centres de documentation dans les écoles primaires. De même, une stagiaire avait travaillé sur le partenariat littéraire privé et associatif à Nantes, en direction de l'édition littéraire. Les résultats obtenus grâce à ces études ont, bien entendu, été inclus dans le travail de formulation des préconisations.

La quatrième phase (novembre à décembre 2009) a correspondu à la phase de mise en débat et d'arbitrage.

Dans tous les cas, la méthodologie que nous retenons s'articule autour de plusieurs volets. Le citoyen se retrouve systématiquement associé, mais ces enquêtes viennent en appui d'autres investigations. Les croisements de points de vue et le partage des résultats nous permettent ensuite d'aller plus loin.

Enjeux et perspectives à l'occasion du nouveau mandat

Les limites de la démarche

Ces évaluations représentent pour la collectivité un processus lourd, qui nécessite un certain volontarisme. Nous nous devons de garantir la rigueur. C'est la raison pour laquelle des moyens importants ont été alloués à la démarche d'évaluation, autour d'un service composé de trois personnes. Le service d'observation sociodémographique renforce nos analyses et le budget qui est alloué à cette mission nous permet de recourir à des prestataires extérieurs. Le Comité opérationnel d'évaluation constitue l'instance d'animation de la démarche, qui assure notamment la veille en matière déontologique (encadré 3).

Les évaluations ne permettent pas généralement d'appréhender l'intégralité d'une politique publique. Elles se concentrent généralement sur des aspects précis. Cependant, brique par brique, les travaux se complètent, permettant de dégager une vision consolidée et plus large de ces politiques et de leurs faiblesses. La mise en perspective des résultats demeure un passage obligé afin de cerner le champ de préconisations opérationnelles.

Il n'est pas toujours simple de trouver des relais au sein de la collectivité pour assurer la mise en œuvre de ces préconisations. En effet, celles-ci intègrent parfois des problématiques de ressources humaines ou de gestion fortes et la Direction se retrouve seule à assurer le portage des conséquences. De fait, parfois, il faut une ou deux années pour parvenir à une mise en œuvre effective.

Il est aussi difficile de capitaliser sur les résultats dans la durée. Nous devons rester particulièrement vigilants pour ne pas questionner à nouveau des éléments qui ont déjà fait l'objet d'enquêtes par le passé. Enfin, les outils et les disciplines évoluent rapidement dans le champ du pilotage des politiques publiques et de leur évaluation. Il nous appartient donc de trouver des articulations avec les autres démarches mises en œuvre, comme la démarche Performance qui s'appuie sur des indicateurs de suivi évaluatif, ou les missions d'information et d'évaluation des assemblées délibérantes.

ENCADRÉ 3

Contraintes et limites de la démarche nantaise

Les contraintes de la démarche :

- Un processus « lourd », adossé à un volontarisme fort
- Quelques moyens
- Un questionnement « limité »
- Une mise en perspective des résultats qui reste relativement cadrée
- Un passage de relais encore balbutiant pour la mise en œuvre des préconisations
- Des résultats difficiles à capitaliser dans la durée
- Des articulations à trouver avec les autres démarches d'évaluation : performance (suivi évaluatif), missions d'information et d'évaluation des assemblées délibérantes des collectivités territoriales (art.8 de la loi du 27/02/2002 relative à la démocratie de proximité)

Et ses perspectives...

L'arrivée d'une nouvelle équipe municipale en mars 2008 a provoqué quelques bouleversements, avec la nomination d'un élu dédié à la démarche d'évaluation. Il s'agit d'un adjoint au maire qui suit aussi les finances et le dialogue citoyen (encadré 4). La Mission évaluation restait, au cours du précédent mandat, portée pour l'essentiel par l'équipe technique. C'est une dimension plus politique qui se met en place aujourd'hui. Les élus seront associés plus étroitement, au niveau de la commande notamment.

Parallèlement, une attention plus grande encore sera portée à l'avenir sur l'expression des besoins et des attentes des citoyens, dans la mesure où la Mission évaluation des politiques publiques se trouve aujourd'hui rapprochée de la Mission dialogue citoyen, dans le cadre du Pôle politiques publiques et prospective. De fait, nos équipes travaillent conjointement. Il est intéressant de noter que la Mission dialogue citoyen relevait autrefois de la communication externe avant d'être intégrée dans les politiques publiques. Ce mouvement traduit en réalité un changement d'attitude. Le dialogue citoyen s'inscrivait auparavant dans une logique d'information. Désormais, il servira à la production de politiques publiques.

Nous voulons que les résultats soient plus largement diffusés auprès du grand public, avec une communication plus large en amont et au cours de ces évaluations. Les contraintes budgétaires devront être mieux prises en compte pour intégrer plus étroitement les notions d'efficacité, par le biais de la démarche Performance, ainsi que par l'intermédiaire du contrôle de gestion. Enfin, nous voulons élargir l'éventail des possibles au moment de la formulation des préconisations, grâce à des méthodes plus prospectives.

ENCADRÉ 4

Un nouveau municipal, de nouveaux enjeux

Les perspectives :

- Un élu dédié, la démarche des ateliers prospectifs
 - Aller vers davantage d'évaluations commanditées par le politique
- Une attention encore davantage portée sur l'expression des attentes et des besoins des citoyens ainsi que dans la mobilisation et la coopération des acteurs sur les enjeux territoriaux
 - Un rapprochement organisationnel avec le dialogue citoyen (vers des évaluations participatives, des méthodologies plus inventives...)
- Des résultats d'évaluation plus largement diffusés qui tombent dans le domaine public
 - Un service ressource en charge de la communication externe
- Des contraintes budgétaires mieux prises en compte (critère de l'efficacité)
 - Des liens à nouer avec la démarche Performance initiée par le contrôle de gestion
- Élargir l'éventail des possibles au moment de la formulation des préconisations
 - Initier les méthodologies inhérentes aux démarches prospectives dans certains contextes ou sur certaines thématiques

L'association des bénéficiaires et des citoyens au processus d'évaluation des politiques régionales

Sylvain Ortis (région Rhône-Alpes)

Collectivité régionale et évaluation

La spécificité de l'action régionale

La collectivité territoriale régionale est composée de deux assemblées : le conseil régional et le Conseil économique et social régional (CESR). Le conseil régional est composé de 157 élus régionaux, élu au suffrage universel direct par départements. La réforme en préparation pourrait modifier le mode de scrutin. Le président de cette assemblée dirige l'exécutif. Le Conseil économique et social régional (CESR) est une assemblée consultative, qui représente la société civile organisée. Ses 116 membres sont associés de très près au processus d'évaluation.

L'administration régionale compte environ 1 100 agents. En intégrant les personnels techniciens, ouvriers et de services (TOS), nous atteignons un total proche de 5 000 agents. Son budget s'élève à 2,47 milliards d'euros. Ses compétences principales résident dans la construction et l'entretien des lycées, la formation continue, l'apprentissage et le transport ferroviaire de voyageurs.

Les régions se distinguent des autres collectivités notamment par leurs modes d'intervention. Trois types de réalisation (encadré 5) peuvent être repérés :

- le soutien financier apporté à des acteurs ;
- le financement du fonctionnement de services publics régionaux confiés à des tiers ;
- la prise en charge par la région de projets en réalisation directe.

Il faut savoir que cette dernière représente le mode d'action le moins répandu pour les régions. Parmi les services publics confiés à des tiers, il faut bien entendu évoquer les trains express régionaux (TER) de la SNCF, tandis que la principale réalisation directe demeure la construction de lycées.

Deux modes de gestion peuvent de même être distingués, d'une part l'attribution de subventions (y compris les dotations) et la gestion directe.

ENCADRÉ 5

Trois types de réalisation ou produits finis

Les modes d'intervention des régions :

1. Soutien financier apporté à des acteurs

- Aides à la création d'entreprises
- Aide à la modernisation de l'hôtellerie rurale
- ...

2. Financement du fonctionnement de services publics régionaux confiés à des tiers

- Fonctionnement du transport ferroviaire des voyageurs (TER)
- Fonctionnement des lycées
- ...

3. Prise en charge par la Région de projets en réalisation directe

- Construction des lycées publics
- ...

Spécificités de l'évaluation des politiques régionales

Ce recours très fréquent à des tiers éloigne la région des bénéficiaires ultimes des politiques et des citoyens, ce qui induit des problèmes particuliers dans l'évaluation.

Comment, en premier lieu, ajuster en connaissance de cause les modalités de mise en œuvre d'une politique publique ? La réponse passe par un suivi systématique et continu, dès l'adoption d'une nouvelle politique, des réalisations de l'action régionale et un bilan périodique de l'exécution des politiques régionales, à destination des élus.

Comment, en second lieu, améliorer les politiques pour qu'elles répondent mieux aux besoins identifiés ? C'est à ce niveau qu'interviennent les évaluations, qui portent en cours de mandature sur les actions lancées au début de celle-ci. Celles-ci sont complétées par des évaluations de fond, qui permettent de réorienter ou de redéfinir des politiques complexes mises en œuvre depuis au moins quatre ans.

La région Rhône-Alpes faisait partie des pionniers en matière d'évaluation, puisque les premiers travaux de ce type ont été entrepris dès 1992. Le dispositif a été revisité à deux reprises, dont la dernière fois en 2005, avec l'arrivée du nouvel exécutif. Celui-ci a voté une délibération concernant l'évaluation et qui fixe une série de principes à respecter.

Les cinq principes de l'évaluation en Rhône-Alpes

- Parmi ceux-ci figure, tout d'abord, la prépondérance du politique.

Les évaluations émanent du président et des vice-présidents. Les mandats d'évaluation sont votés par le Conseil et un débat sur la démarche généralisée d'évaluation se tient chaque année. Au cours de ce débat, toutes les évaluations conduites sont présentées à l'Assemblée et les conclusions et recommandations sont discutées. Chaque étape (avant, pendant et après l'évaluation) donne lieu à un débat dans les commissions thématiques spécialisées, qui rassemblent une vingtaine d'élus régionaux. Cette prépondérance du politique trouve sa logique dans le fait que c'est l'exécutif qui décide des évolutions à apporter.

- L'impartialité constitue la deuxième exigence.

Tous les travaux d'évaluation sont réalisés par des prestataires externes. Le pilotage est confié à un groupe *ad hoc*, qui est présidé par une personnalité extérieure à l'institution. Il s'agit donc d'une personne qui connaît la politique évaluée, mais qui n'a pas d'intérêt direct dans celle-ci. Ce ne sont donc pas les vice-présidents en charge de la politique publique évaluée qui président le comité de pilotage, comme c'est le cas dans d'autres collectivités. Enfin, chaque rapport d'évaluation est soumis à l'avis d'experts en évaluation de politiques publiques.

- La diffusion de la culture d'évaluation et des résultats des travaux constitue une troisième exigence.

Les élus et les services sont ainsi amenés à suivre des formations et la Région participe à des colloques et manifestations diverses.

- L'ouverture à la société civile constitue la quatrième exigence.

Celle-ci se traduit par l'intégration d'une dizaine d'opérateurs et de bénéficiaires dans chaque comité de pilotage. La démarche d'évaluation doit s'articuler avec les principes de la démocratie participative. La diffusion des résultats et l'association du CESR (Conseil économique et social régional), consulté en amont de l'évaluation et inclus dans le comité de pilotage, s'inscrivent dans le cadre de ce quatrième principe.

- Enfin, le dernier principe a trait à la cohérence d'ensemble de cette démarche.

Le pilotage de la démarche est confié à un Comité de suivi de la démarche d'évaluation composé de deux collèges (les représentants des groupes politiques du Conseil et des experts en évaluation des politiques publiques). L'évaluation doit aussi s'articuler avec d'autres démarches transversales :

- développement durable ;
- territorialisation des politiques ;
- juste représentation des sexes, des âges, des catégories socioprofessionnelles ;
- complémentarité avec le comité de suivi et d'évaluation des fonds accordés aux entreprises.

Association des bénéficiaires aux évaluations des politiques régionales

Les bénéficiaires sont associés au comité de pilotage, qui constitue un groupe ressources pour l'évaluateur, dans la mesure où il rassemble toutes les parties prenantes de la politique évaluée. Ils sont aussi associés aux investigations, avec pour objectif de recueillir leur point de vue.

L'exemple de l'évaluation de la carte M'ra !

Le cas de l'évaluation de la carte M'ra ! illustre bien cette politique d'association des bénéficiaires (encadré 6). La carte M'ra ! est distribuée gratuitement à tous les lycéens, apprentis et jeunes dépendant des missions locales. Elle offre un certain nombre d'avantages dans les domaines du sport, de la culture et l'achat des manuels scolaires. Quelque 5 000 partenaires ont été conventionnés et l'on dénombre 300 000 bénéficiaires (encadré 7). La distribution de cette carte vise à concourir à l'égalité des chances, à développer l'accès à un bon niveau de formation et à la culture ainsi qu'à favoriser la pratique sportive (schéma 6).

ENCADRÉ 6

Le fonctionnement de la carte M'ra

- Modalités d'action de la Région
 - Aide à l'achat de manuels scolaires
 - Aide pour l'achat d'un livre loisir
 - Financement d'un abonnement de 6 places de cinéma d'art et d'essai
 - Aide à l'achat d'un abonnement et/ou de places d'entrée dans tous les sites culturels de spectacle vivant et festivals partenaires
 - Gratuité d'accès dans les lieux d'art et de savoirs de la région (musées, sites patrimoniaux, centres d'art, lieux de mémoire...)
 - Aide à la prise d'une licence sportive auprès des instances sportives fédérales ayant signé une convention cadre avec la Région
- Fonctionnement
 - Remise aux bénéficiaires d'une carte à puce pluriannuelle et gratuite
 - Carte utilisable dans les librairies, les associations de parents d'élèves et l'ensemble lieux culturels et sportifs conventionnés avec la région
 - Système de paiement installé chez les partenaires agréés par la région (libraires, associations, salles de cinéma, de spectacle, musée...)
- Publics éligibles
 - Pour les manuels scolaires : élèves et apprentis scolarisés dans un établissement de formation initiale relevant de la compétence de la région ou préparant un diplôme, une formation complémentaire d'initiative locale, ou un titre homologué de niveau V ou IV, ou encore fréquentant une classe de 3^e découverte professionnelle/insertion, 3^e ou 4^e technologique, une section d'enseignement général et professionnel
 - Publics complémentaires pour les avantages culture et sports : jeunes inscrits en BTS, classes préparatoires et formations post-bac intégrées, dans des structures pour la jeunesse handicapée (IME, IMPRO, IMP), des structures de placement en hébergement (avec inscription dans un cursus de formation), des missions locales et PAIO.

ENCADRÉ 7

La carte M'Ra en chiffres**2005-2006 :****295 000 bénéficiaires**

- Environ 16,5 millions d'euros par an pour les manuels scolaires
- Environ 3,4 millions d'euros par an pour le volet culturel
- Environ 1,3 million d'euros par an pour le volet sport

Montant maximum d'avantages mobilisables

par carte : 198 euros (hors gratuité d'accès dans les lieux d'art et de savoirs)

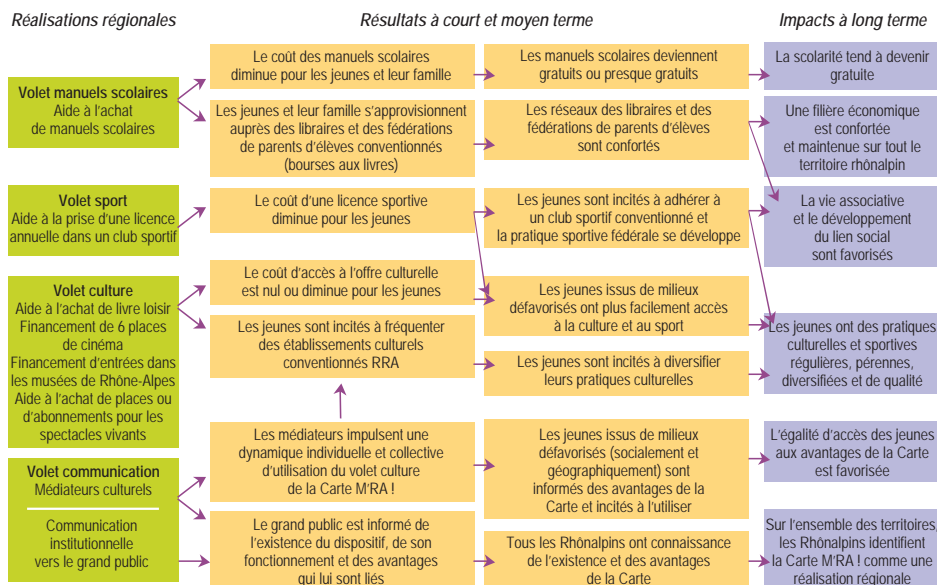
- Montant moyen par jeune éligible aux manuels scolaires : 73
- Montant moyen par jeune hors manuels scolaires : 17

Taux de transformation des avantages

- Manuels scolaires : 85 %
- Livre loisir : 40 %
- Cinéma : 36 %
- Spectacle vivant : 24 %

SCHEMA 6

Synthèse de la logique d'action pour la carte M'ra !



Pour la région, le coût s'avère assez important, aussi bien au niveau des prestations que de la gestion. Il a donc été décidé d'en évaluer les résultats.

Le comité de pilotage se compose de :

- un président (inspectrice d'académie en retraite) ;
- sept représentants des groupes politiques du conseil régional ;
- deux représentants du CESR ;
- deux citoyens non bénéficiaires ;
- quatorze bénéficiaires directs et indirects de la politique.

Ces mêmes acteurs se retrouvent dans la phase de recueil de données. Plusieurs méthodologies ont été mises en œuvre : des entretiens en face-à-face avec des opérateurs et des bénéficiaires, des entretiens téléphoniques et des enquêtes en ligne à destination des anciens bénéficiaires. Cet accent mis sur les anciens bénéficiaires avait pour ambition de déterminer si cette carte avait induit de nouvelles pratiques. Par ailleurs, des réunions de travail ont été organisées avec le Conseil académique de la vie lycéenne pour travailler sur les recommandations, de même que des *focus groups* dans les lycées. Nous préférons en effet que les préconisations soient discutées ouvertement par les acteurs, plutôt qu'elles ne soient formulées et votées directement en chambre.

Association des citoyens aux évaluations de politiques régionales

Les bénéficiaires ne sont pas les seuls intéressés

L'association des citoyens aux évaluations des politiques régionales a été initiée en juin 2005 mais beaucoup reste à accomplir. Ainsi que le remarquait Jean-Jack Queyranne, président du

conseil régional, à l'occasion des Journées françaises de l'évaluation organisées à Lyon en 2006 : « *Le citoyen est toujours le destinataire final des politiques publiques, mais il est de plus souvent électeur et contribuable, ce qui fait de lui aussi, en quelque sorte, le commanditaire et le financeur de ces politiques. Il a donc toute légitimité à être associé aux démarches d'évaluation.* »

Nous estimons ainsi que les bénéficiaires ne sont pas les seuls intéressés par une politique publique. Prenons l'exemple de l'enseignement supérieur. Les étudiants ne doivent pas être les seuls à être consultés. Les parents, les anciens étudiants, les citoyens engagés doivent pouvoir faire valoir leur point de vue. Au niveau d'une grande région, il est difficile d'interroger tous les citoyens individuellement. Nous voulons tout de même les associer largement, avant, pendant et après l'évaluation. C'est pourquoi nous avons convié en 2006 un panel de 30 citoyens à se pencher sur cette question.

L'« expérimentation citoyenne » et ses conclusions

Ce groupe a produit un avis remis au Conseil régional et à la Société française d'évaluation. Il dessine des pistes multiples sur les attentes du public quant à la manière de l'intégrer dans l'évaluation. L'expérimentation citoyenne, qui a trouvé sa concrétisation dans la réunion de ce groupe de citoyens, s'est déclinée suivant cinq phases (encadré 8). La première a consisté en la sélection des citoyens, qui ont dû être ensuite sensibilisés sur les principales notions d'évaluation des politiques publiques. Ils ont ensuite participé aux travaux des 7^{es} journées françaises de l'évaluation. Un certain nombre de réunions de *debriefing* ont ensuite eu lieu, avant la production de l'avis en juillet 2006.

Le travail a été organisé autour de trois grandes réflexions :

- Quelle est la place du citoyen dans l'évaluation ? Quel intérêt à voir son rôle se renforcer ?
- Quel pourrait être le rôle des citoyens tout au long du processus d'évaluation ?
- Quelles sont les conditions à remplir pour que les citoyens puissent occuper la place souhaitée ?

Dans leur rapport, les membres du groupe déclarent : « *Nous pensons donc que le rôle des citoyens dans l'évaluation n'est pas d'exercer un contre-pouvoir par rapport à ceux qui détiennent le pouvoir ; il est d'apporter un point de vue supplémentaire dans l'évaluation, au même titre que celui des élus, des gestionnaires, des experts et des autres acteurs, pour enrichir le processus de décision ultérieur.* » Pour mettre en œuvre cette recommandation, nous avons décidé entre autres d'associer des citoyens aux comités de pilotage.

Par ailleurs, les auteurs notent : « *Une évaluation plus citoyenne ne doit pas devenir une somme d'intérêts particuliers. Chaque citoyen pris isolément n'a qu'un angle de vue spécifique et ce n'est que « l'intelligence collective » des citoyens qui fonde une vision plus complète et plus complémentaire des politiques publiques.* »

Peut-être plus importante encore est la conclusion de ce document : « *Nous voudrions indiquer que nous ne proposons pas de mettre en place une "évaluation citoyenne" à côté d'une "évaluation d'experts". Nous voulons même bannir ce terme d'"évaluation citoyenne". Ce que nous voulons, c'est participer à l'évaluation des politiques publiques qui doit rester un seul processus, enrichi par la participation des citoyens.* » Ils conçoivent ainsi l'évaluation comme un ensemble cohérent et réfutent l'idée d'une évaluation parallèle effectuée par les citoyens. C'est le parti pris que nous avons retenu, en écartant notamment l'idée des jurys citoyens.

ENCADRÉ 8

L'expérimentation citoyenne: modalités de mise en œuvre**Objectif de l'expérimentation**

Permettre l'expression d'un point de vue citoyen sur l'évaluation des politiques publiques

Mise en œuvre de l'expérimentation en 5 phases

1. Sélection des citoyens pour participer à l'expérimentation (avril et mai 2006)
2. Sensibilisation du groupe citoyen aux principales notions d'évaluation des politiques publiques
3. Participation du groupe aux travaux des 7^{es} Journées françaises de l'évaluation (JFE)
4. Réunions de débriefing après les 7^{es} JFE
5. Production d'un avis citoyen relatif à l'évaluation des politiques publiques (juillet 2006)

Lors des JFE, les citoyens ont participé à l'ensemble des tables rondes, des ateliers et agoras au même titre que les autres congressistes

Le groupe a structuré sa réflexion autour de 3 grandes interrogations

- Quelle est la place du citoyen dans l'évaluation ?
- Quel intérêt à voir ce rôle renforcé ?
- Quel pourrait être le rôle des citoyens tout au long du processus d'évaluation ?
- Quelles sont les conditions à remplir pour que les citoyens puissent occuper la place souhaitée ?

La méthode retenue pour associer les citoyens

De ces conclusions, il est ressorti qu'il était possible d'impliquer les citoyens dans le choix des thèmes d'évaluation. Cette option n'a cependant pas été retenue, car nous considérons qu'il s'agit d'une décision politique, à la main de l'exécutif. De plus, seuls les grands dispositifs ont vocation à être évalués.

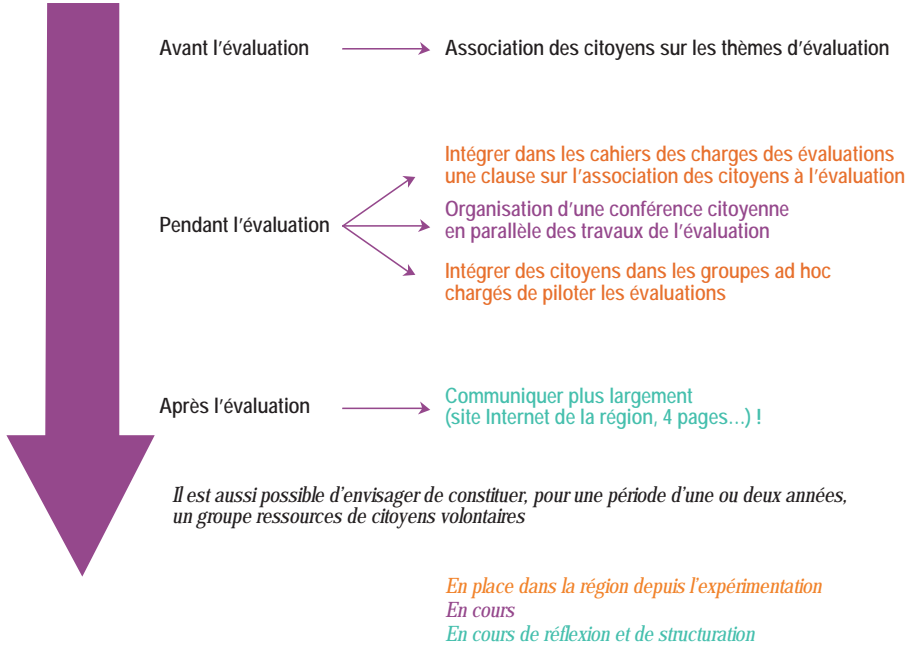
Au cours de la deuxième étape, celle de l'évaluation proprement dite, il a été décidé d'intégrer dans les cahiers des charges une clause sur l'association des citoyens (schéma 7). Le prestataire doit donc intégrer dans sa méthodologie le recueil du point de vue des citoyens. Généralement, cela se traduit par une enquête par téléphone. Pour l'heure, nous restons quelque peu déçus par ce manque d'originalité et d'initiatives. Nous avons aussi décidé de mettre en place des conférences citoyennes en parallèle de l'évaluation, et nous avons intégré depuis 2006 des citoyens dans les comités de pilotage *ad hoc*. D'un côté, ils apportent une vision différente et participent pleinement aux débats. De l'autre, au milieu d'une trentaine d'experts, il n'est pas simple de s'imposer. Pour éviter cela, ils sont réunis avant la première séance du comité pour mieux comprendre la politique, la logique d'action et les buts de l'évaluation.

Enfin, un travail a été entrepris sur la communication des résultats des évaluations. Les rapports sont tous disponibles, mais ils restent peu accessibles au grand public. C'est pourquoi nous essayons de les rendre moins complexes et de développer des outils de communication appropriés.

SCHÉMA 7

Quand et comment associer les citoyens à l'évaluation ?

Les orientations retenues



La place des citoyens dans l'évaluation des politiques régionales a ainsi été introduite tout d'abord avec la constitution en 2006 du groupe composé de trente citoyens (comme on l'a vu plus haut). Il s'agissait là d'un premier pas vers une association plus étroite des citoyens, au même titre que leur inclusion dans les groupes *ad hoc*. Parmi les axes de développement, on peut citer la réflexion menée sur la communication des résultats de l'évaluation et l'évaluation en cours de la politique régionale en faveur de la démocratie participative.

Les enseignements de l'expérience rhônalpine

Nous avons véritablement ressenti un appétit citoyen pour l'évaluation des politiques publiques, comme l'a démontré l'expérimentation citoyenne de 2006. Les citoyens qui font partie des comités de pilotage sont pour l'instant choisis de manière systématique parmi les 30 membres du groupe de l'expérimentation citoyenne, car ils sont déjà sensibilisés aux questions d'évaluation.

La question de la représentativité des citoyens dans les instances a été largement abordée à cette occasion. La représentativité est déjà une réalité au sein de certaines instances comme le CESR. Cependant, au niveau de la représentation des citoyens organisés, nous constatons d'une part une certaine dérive lobbyiste et d'autre part d'une dérive *intuitu personae*, qui pousse certains à représenter leur personne avant de représenter leur groupe. De plus, nous observons que certaines catégories sont surreprésentées par rapport à d'autres. Au niveau de la représentation des citoyens non organisés, le risque de dérive *intuitu personae* est plus évident encore. Des questions se posent également concernant la constitution des échantillons.

Les citoyens rhônalpins ont montré un appétit certain et veulent être informés des résultats de l'évaluation (encadré 9). Cependant, la complexité des sujets rend difficile la communication. Tout reste à faire sur ce plan pour rendre lisibles et complets les rapports d'évaluation.

ENCADRÉ 9

Les enseignements de l'expérience rhônalpine

Appétit citoyen pour les évaluations de politiques publiques

Succès de l'expérimentation citoyenne et des appels à candidatures pour participer aux différents groupes *ad hoc* constitués lors des évaluations
Pré-requis nécessaires quant à la connaissance des politiques évaluées
Une séance d'information et de formation spécifique

Représentativité citoyenne : un idéal hors de portée ?

La représentativité des citoyens organisés (société civile organisée)
Dérive lobbyiste et dérive *intuitu personae*?
Catégorie surreprésentée au détriment d'autres (demandeurs d'emplois...)

La représentativité des citoyens non organisés

Modalités de constitution d'échantillons de citoyens pour les tirages au sort
Risque de l'*intuitu personae*?

Communication et diffusion des résultats des évaluations : tout reste à faire !

La transparence à l'épreuve du débat politique
Des sujets abrupts et complexes à rendre lisibles et concrets

L'exemple de l'évaluation de la politique en faveur de la démocratie participative

Pour terminer, je souhaiterais évoquer la dernière évaluation en cours, autour de la politique en faveur de la démocratie participative. Il nous paraissait tout naturel d'associer très étroitement les citoyens à cette évaluation concernant une politique lancée en 2005. Elle vise à subventionner des porteurs de projets (associations et autres collectivités territoriales) et à organiser des ateliers citoyens sur diverses thématiques. Ces derniers ont pesé de manière remarquable dans la construction de nouvelles politiques publiques. Un réseau régional de démocratie participative a été mis en place, avec 250 acteurs qui se réunissent régulièrement pour échanger sur leurs pratiques autour de cette question. Des rencontres multiples ont en outre été organisées dans ce cadre.

Trois grandes questions évaluatives se posent :

- Dans quelle mesure la démarche régionale a-t-elle permis de renforcer l'intérêt des citoyens pour la vie politique et l'action publique ?
- En quoi la politique régionale de démocratie participative a-t-elle modifié les politiques et les pratiques de la Région et de ses partenaires ?
- Quel est l'effet du soutien régional auprès des porteurs de projets et de leurs publics cibles ?

De manière très originale, un atelier de réflexion a été mis en place pour enrichir le travail d'évaluation. Il se compose de deux groupes avec, d'une part, vingt citoyens rhônalpins tirés au sort et, d'autre part, 20 élus du territoire régional. Il est à noter à ce propos qu'il nous a fallu appeler 2 500 personnes pour trouver les 20 citoyens qui acceptent de participer à cet atelier. Les contraintes imposées aux personnes sélectionnées étaient en effet assez lourdes, puisqu'elles devaient se rendre disponibles trois week-ends d'affilée. Elles

ont travaillé autour de la question : « Associer les citoyens à la politique régionale : quel sens cela a-t-il ? » Il ne s'agit donc pas d'une question évaluative, mais d'une question générale et prospective. Ce travail servira essentiellement à la formulation de recommandations. Le prestataire juge l'avis des citoyens particulièrement intéressant. Il est probable que nous renouvellerons cette expérience dans le futur, dans le cadre d'autres évaluations.

SYNTHÈSE DES DÉBATS AVEC LA SALLE

Le choix des évaluations

Questions de la salle

À Nantes, comme au niveau de la région Rhône-Alpes, le choix des évaluations est laissé à l'exécutif. Vous écarterez l'initiative citoyenne en la matière. Cependant, une différence apparaît entre les deux approches. Il semblerait que la ville de Nantes évalue une politique avec une idée précise de l'axe à améliorer. À Lyon, l'évaluation sert essentiellement à vérifier le bon fonctionnement d'une politique et corriger d'éventuels défauts, sans *a priori* sur leur nature. Quels sont en fait les éléments dont disposent les élus avant de décider d'une évaluation? Comment choisissent-ils les thèmes des évaluations? S'agit-il d'un choix uniquement politique ou orienté par des études préexistantes?

Réponse de Mme Fenet

Il est vrai qu'à la ville de Nantes, les évaluations sont souvent lancées à partir d'un constat d'insatisfaction. Ce sont des pétitions d'usagers ou des rencontres avec les citoyens qui permettent d'alerter sur un problème particulier. L'élu est donc directement interpellé et peut déterminer s'il est nécessaire de lancer une démarche d'évaluation. Cette dernière va permettre d'objectiver les difficultés et de détecter les éventuelles pressions provenant de lobbies. Parfois, les conclusions de l'évaluation sont à l'opposé des *a priori* que nous avions avant le lancement de l'évaluation. Ainsi, nous pensions au départ que les assistantes maternelles étaient mal implantées géographiquement et que les horaires n'étaient pas adéquats. Ces éléments ne se sont pas retrouvés dans le diagnostic après évaluation.

Souvent, les bons diagnostics sur la situation ne suffisent pas pour déterminer les bonnes solutions. Par exemple, dans le cas du maintien des personnes âgées à domicile, nous ne savions pas *a priori* comment répondre aux manques qui semblaient évidents en la matière. Le travail d'évaluation a permis d'élaborer des réponses opérationnelles, dans les domaines de l'adaptation des logements sociaux, de la lutte contre l'isolement et de l'aide aux aidants.

Réponse de M. Ortis

Le Conseil régional ne prend pas en compte les demandes des populations en matière d'évaluation, mais les associe étroitement au processus jusqu'au moment de la construction de la politique.

Il est clair que les Directions opérationnelles ne se lancent pas dans une évaluation sans *a priori*. Cela semble bien naturel et fort sain. L'évaluation aura le mérite d'objectiver et de découvrir de nouveaux éléments. Le fait qu'un élément externe à l'administration mette ses conclusions par écrit prend une valeur importante.

Le contrôle de gestion fournit des éléments importants sur les réalisations plutôt que sur les résultats. C'est la mise en œuvre qui se retrouve au cœur du débat dans le cadre de cette démarche. Les éléments qui en ressortent n'indiquent en rien si les objectifs ont été atteints. Cependant, ils s'avèrent tout de même ensuite très utiles à l'évaluateur, notamment les bilans physico-financiers.

Complément de Mme Daniel

La démarche Performance instaure un système de veille permanent, qui a notamment vocation à mettre en évidence des difficultés, qui peuvent ensuite donner lieu à une évaluation. Le travail visant à mettre au point une batterie d'indicateurs a débuté en 2006. Nous n'avons pas encore le recul suffisant pour faire en sorte que ce système puisse concourir au déclenchement d'évaluations. Il est probable que cela ne soit pas véritablement le cas à terme.

Dans le sens inverse, il était prévu que la Mission évaluation concoure à la mise au point d'indicateurs pour le suivi évaluatif. Ce travail semble en meilleure voie. Dans le cadre de l'évaluation sur la politique de lecture, le contrôle de gestion a été associé pour déterminer si les résultats recueillis peuvent alimenter ce service.

Le champ de compétence des évaluations

Question de la salle

Souvent, les besoins des citoyens débordent du champ de compétence d'une seule collectivité. Comment résoudre le dilemme qui se pose dès lors ?

Réponse de Mme Fenet

Pour contourner ce problème, il apparaît essentiel d'associer les autres collectivités et partenaires qui disposent de leviers sur une politique publique donnée. En effet, bien souvent, la politique publique menée sur un territoire est la résultante de la mobilisation de divers niveaux et acteurs. L'élaboration du cahier des charges et le pilotage de la démarche sont réalisés en relation étroite avec ces acteurs. Ainsi émerge un plan d'actions partagé. En ce qui concerne les assistantes maternelles, la Ville agit essentiellement par le biais du RAM, alors que le Conseil général dispose d'une compétence forte en la matière. Beaucoup de recommandations s'adressaient essentiellement à ce niveau. Nous menons en ce moment une évaluation sur les modes de garde et les freins à l'emploi. Pour celle-ci, la mission locale et la Communauté urbaine sont étroitement associées. Bien entendu, cette imbrication de différents niveaux rend le pilotage plus délicat.

La méthode d'évaluation

Questions de la salle

La palette classique des outils se retrouve de part et d'autre, autour d'enquêtes, d'études statistiques, d'interrogations multiples et de réunions avec des partenaires. Après vos exposés se pose la question de l'externalisation et de son opportunité. Bien souvent, les prestataires extérieurs n'ont pas de méthode prédéfinie en vue de l'analyse des données qualitatives ou quantitatives. Pour votre part, avez-vous arrêté des consignes claires en la matière dans le choix de ces prestataires ?

Réponse de M. Ortis

Les évaluations sont externalisées et le Conseil régional n'est que commanditaire. Ceci permet de nous dégager des pressions administratives et politiques. Nous partons du principe

que les conclusions ne s'imposent pas à l'exécutif. L'évaluation n'est perçue que comme une aide à la décision. De fait, de nombreuses recommandations ne sont jamais mises en œuvre, mais nous ne censurons en aucun cas le prestataire. L'existence d'un service dédié à l'évaluation constitue un garde-fou pour nous prémunir des pressions. Ce service joue un rôle de tampon entre la Direction opérationnelle et le prestataire. Nous nous assurons ainsi que le travail sera mené de la façon la plus objective et impartiale possible. Le fait d'associer les élus de l'opposition à toutes les étapes (vote du mandat et participation au comité de pilotage) constitue une autre garantie, même si au final, les décisions appartiennent à l'exécutif. De même, la démarche d'ensemble est pilotée par un comité de suivi de la démarche de l'évaluation, présidé par un élu de la majorité, mais il compte parmi ses membres un élu de chaque groupe représenté au Conseil.

Nous sommes confrontés aux problèmes que vous soulevez en matière de construction et d'analyse des données. À ce propos, au moment du choix du prestataire, nous privilégions les groupements d'entreprises face aux professionnels de l'évaluation. Souvent, les évaluateurs professionnels se retrouvent ainsi contraints de s'adjoindre les services d'un cabinet spécialisé dans le domaine qui est évalué.

Réponse de Mme Daniel

En tant que mission d'évaluation, nous nous fixons pour objectif de garantir la neutralité vis-à-vis des prestataires externes. Ceci n'est pas simple, car il s'agit d'acteurs qui collaborent de manière fréquente avec la Ville et qui nouent des relations avec ses représentants. Pour que l'analyse soit la plus distanciée possible, nous n'hésitons pas à inclure des éléments omis par le prestataire. C'est ensemble que les acteurs préservent une certaine distance vis-à-vis du politique.

Je voudrais maintenant revenir sur la question du choix des méthodes par les prestataires. Nous sommes très exigeants sur les méthodologies au moment de la sélection du prestataire. Nous restons de même très vigilants au moment de la remise des conclusions. Nous avons renforcé la compétence « études » de la Mission évaluation, en ce qui concerne les études qualitatives comme les études quantitatives. Par ailleurs, nos experts extérieurs, en provenance du CNRS pour l'essentiel, disposent de compétences propres. Nous savons nous montrer très pointilleux en la matière pour garantir une certaine rigueur. Cependant, nous commençons à bien connaître les prestataires et les CV des intervenants. Nous en discutons ensemble au niveau de la mission, sans mettre toutefois ces questions en débat au sein du comité de pilotage, car cette instance s'intéresse peu à ce type de problèmes.

La participation des citoyens

Question de la salle

Nous cherchions aujourd'hui à savoir comment intéresser les citoyens au processus évaluatif et à l'apport de celui-ci dans ce cadre. Ce sujet n'a pas été pleinement exploré dans vos présentations. Qu'est-ce qu'un citoyen et comment le faire participer réellement ? Sylvain Ortis, vous avez évoqué les difficultés que vous rencontriez pour recruter des citoyens. Il vous a ainsi fallu appeler plus de 2 000 personnes pour recruter un groupe de vingt volontaires. Je me demande cependant si vous avez adopté une méthode de sélection particulière, afin d'obtenir un panel représentatif. En effet, le hasard risque d'introduire des biais dans les réponses de l'instance.

Réponse de Mme Daniel

Il faut distinguer plusieurs modes d'associations des citoyens. D'une part, il est possible de recueillir les points de vue, par le biais d'enquêtes notamment. Dans ce cadre, il est important de préserver la représentativité. D'autre part, nous pouvons aussi associer les citoyens au pilotage de l'évaluation. À ce niveau, il semble difficile de constituer des échantillons représentatifs. C'est plutôt le principe de la diversité que nous retenons alors. Nous avons testé cette modalité sur la question du logement des jeunes. Nous avons sélectionné des jeunes qui faisaient partie du Conseil nantais de la jeunesse. Ainsi, nous savions que nous allions inclure des personnes engagées qui souhaitaient que leur voix soit entendue. Ils apportent une distance supplémentaire dans nos questionnements et renforcent la pertinence de la démarche.

Réponse de M. Ortis

Sans être statisticien, j'ai beaucoup de mal à croire qu'un groupe de vingt citoyens puisse être réellement représentatif de 6 millions de personnes. Nous ne recherchons pas en fait la représentativité, mais la diversité des profils. Pour la constitution de ce groupe, la prestation a été externalisée et confiée à Ipsos, car nous ne disposions pas des compétences en la matière. Nous avons cependant édicté des critères, afin d'assurer la mixité concernant le département d'origine, la catégorie socioprofessionnelle, le sexe et l'âge.

La diffusion des résultats

Question de la salle

Dans quelle mesure l'évaluation peut-elle servir directement au citoyen ? Vous inscrivez cet élément dans une logique managériale, d'aide à la prise de décision, mais dans quelle mesure permet-elle de rendre compte ? Il s'agit là en effet d'une mission de toute première importance. Il ne suffit pas d'intégrer le citoyen à la démarche, mais il est important de diffuser les conclusions largement, de les reprendre dans le discours politique et d'en faire la base du débat démocratique. Or j'ai l'impression que cette dimension reste fort théorique. J'aimerais donc savoir si vos collectivités ont su développer cet aspect, et plus particulièrement le Conseil régional Rhône-Alpes, qui prépare les prochaines échéances électorales.

Réponse de M. Ortis

J'estime que le Conseil régional est très mauvais en matière de diffusion des résultats des évaluations. Les citoyens ne sont pas directement concernés par les résultats de l'évaluation. En revanche, cette démarche entre pleinement dans la logique d'aide à la décision des élus régionaux. Une session du Conseil est consacrée chaque année à l'étude des conclusions de toutes les évaluations. Les vice-présidents concernés annoncent ce qu'ils retiennent de l'évaluation et déclinent les recommandations qu'ils s'approprient à adopter. Un an plus tard, ils détaillent les actions effectivement mises en place. Cela participe du débat démocratique. Cependant, aucun citoyen rhônalpin ne nous a interpellés sur les conclusions des évaluations, pas plus que les médias.

Nous entrons maintenant dans une période électorale. Peut-être les résultats des évaluations vont-ils être repris. En tout cas, ils n'ont jamais été cachés. Les élus de l'opposition comme de la majorité s'en saisissent dès leur publication.

Réponse de Mme Fenet

Les élus attendent avant toute chose de pouvoir témoigner de l'évolution du service rendu à la population. L'une des premières évaluations réalisées par la Ville portait sur les horaires d'accueil périscolaire. Les élus avaient pour première ambition d'élargir les plages horaires et d'améliorer la qualité d'accueil.

La communication directe avec les citoyens est plus simple pour un élu municipal que pour un élu régional, dans la mesure où les villes travaillent beaucoup sur des services à la vie quotidienne, ce qui permet d'engager un débat plus aisément. Les documents recto verso qui reprennent les conclusions des évaluations sont diffusés auprès du public. La revue municipale se fait de plus en plus l'écho du lancement d'évaluations. Par ailleurs, la mise à disposition des questionnaires dans les lieux publics attise la curiosité des citoyens. Des conférences de presse ont été organisées pour présenter les conclusions et les pistes d'amélioration des services. L'élu n'a intérêt à rendre les résultats publics que dans cette seule perspective.

À la fin du précédent mandat ont été lancées plusieurs évaluations destinées à soutenir le travail des futurs élus. Il serait étonnant que les élus candidats mobilisent les productions mises en œuvre par les services municipaux. Ces éléments sont disponibles. Il appartient donc aussi au citoyen ou à l'opposition de s'en saisir.

BIBLIOGRAPHIE

- Daniel A.**, 2008, *Évaluation de la politique municipale de la lecture : quelle complémentarité de l'offre pour un égal accès de tous à lecture ?*, Document de travail du pôle « Politiques publiques et prospective » de la Ville de Nantes.
- Ortis S.**, 2009, *Implications des citoyens dans les évaluations de politique de la région Rhône-Alpes*, 9^{es} Journées françaises de l'évaluation, Marseille.
- Région Rhône-Alpes**, 2006, *Expérimentation citoyenne : Avis du groupe de citoyens relatif à la place des citoyens dans l'évaluation des politiques publiques*, 7^{es} Journées françaises de l'évaluation, Lyon.
- Ville de Nantes**, 2007, *L'accueil de la petite enfance par les assistantes maternelles agréées indépendantes : synthèse des résultats de l'évaluation – volet 2*, publication de la Direction générale prospective, évaluation et projets de la ville de Nantes.

L'APPROCHE OBJECTIFS/INDICATEURS

La cinquième séance du séminaire sur les méthodes d'évaluation des politiques publiques s'est intéressée à la démarche qui articule la définition d'objectifs pour la conduite des politiques publiques et la mise en place d'indicateurs permettant de mesurer l'atteinte de ces objectifs.

– Jean-René Brunetière (Conseil général de l'environnement et du développement durable) a ainsi présenté dans un premier temps la conception et l'utilisation des indicateurs dans le cadre de la Loi organique relative aux lois de finances (LOLF).

– Dans un deuxième temps, Isabelle Engsted-Maquet (Direction générale « Emploi, affaires sociales et égalité des chances ») a présenté les exercices de *benchmarking* réalisés par la Commission européenne dans le cadre de la méthode ouverte de coordination (MOC). Ces deux interventions ont été suivies d'un débat avec la salle.

Les indicateurs de la LOLF

Une occasion de débat démocratique ?

Jean-René Brunetiere (CGEDD)

La loi organique relative aux lois de finances (LOLF) de 2001 a imposé que l'État se dote à partir de 2006 d'objectifs et d'indicateurs capables de mesurer les résultats des politiques publiques. Celles-ci sont classées en missions et programmes. En principe, une mission correspond à une politique publique au sens large et un programme à la partie de cette politique publique qui concerne un ministère donné. En 2001, étaient comptabilisées 34 missions et 130 programmes, ainsi que 1 300 indicateurs correspondant à 600 objectifs.

L'analyse des indicateurs LOLF élaborés en 2006

Échantillonnage et méthode d'analyse

L'association Pénombre à laquelle j'appartiens s'intéresse et s'alarme parfois de l'usage des nombres dans les médias, les débats politiques et le débat social. À ce titre, la Revue française d'administration publique (RFAP) m'a demandé de rédiger un article sur les indicateurs dans le cadre de son numéro spécial sur la LOLF en 2006, première année d'application de la loi. Afin qu'il ne me soit pas reproché d'avoir sélectionné trop peu d'indicateurs, je me suis basé sur un échantillon de 32 missions en sélectionnant l'indicateur 1 pour la première mission, l'indicateur 2 pour la deuxième mission et ainsi de suite, les missions étant classées par ordre alphabétique et les indicateurs dans l'ordre dans lequel ils apparaissent dans le texte du ministère des Finances. J'ai cherché à déterminer dans mon analyse si ces indicateurs reflétaient les objectifs correspondants, s'ils mesuraient un résultat de l'action de l'État, s'ils étaient utiles et s'ils n'entraînaient pas d'effets pervers.

Les défauts des indicateurs LOLF

Les indicateurs peu significatifs

Pour les indicateurs peu significatifs, on peut citer celui correspondant à la première mission qui concerne les Affaires étrangères. Son objectif vise à défendre les intérêts de la France à l'étranger. L'indicateur choisi est le nombre de consultations des sites Internet du ministère

des Affaires étrangères et des postes. Or selon cet indicateur, seul le webmestre contribuerait au rayonnement de la France à l'étranger, et non le corps diplomatique.

D'autres cas similaires existent, notamment quand un opérateur réalise l'essentiel de la politique publique comme dans la branche de l'Assurance maladie ou dans les hôpitaux.

Les indicateurs distordus et distordants

Parmi les indicateurs distordus et distordants, on trouve le taux de réclamations contentieuses relatives à la taxe d'habitation traitées dans le délai d'un mois censé mesurer l'activité du service qui traite les réclamations contentieuses relatives à plusieurs impôts. Cet indicateur qui a été choisi dans le but de ne pas multiplier les indicateurs pour chaque impôt risque en effet d'inciter les fonctionnaires de ce service à traiter prioritairement les réclamations contentieuses relatives à la taxe d'habitation au détriment des autres réclamations. L'action publique en est alors distordue.

Ainsi, dans les services du ministère de l'Équipement, le permis de construire devient tacite si l'administration n'a pas traité la demande dans un délai de deux mois. En revanche, aucune disposition n'existe si le délai de deux mois est dépassé dans le cas des demandes de certificat d'urbanisme. En conséquence, dans un service particulier, 53 % de demandeurs de certificats d'urbanisme sont mécontents contre seulement 12 % de demandeurs de permis de construire.

Le deuxième cas d'indicateur distordu concerne le nombre d'affaires pénales qui sont traitées par un magistrat à l'instruction dans les tribunaux de grande instance. Dans ce cas, seule la quantité est mesurée au détriment de la qualité. Si les magistrats considéraient cet indicateur avec sérieux, ils classeraient le maximum de dossiers, en commençant par les plus faciles. En fait, la cible à atteindre pour chaque magistrat s'élève à 61 dossiers par an, ce qui correspond à la moyenne réelle.

Troisième exemple, la part de crédits de formation des professeurs des écoles consacrée à la spécialisation des personnels d'adaptation et d'intégration scolaire (AIS) qui mesure leur part de formation consacrée à la manière d'enseigner aux enfants handicapés. Or pour faire évoluer un quotient, il suffit d'augmenter son numérateur ou de diminuer son dénominateur. En l'occurrence, la deuxième solution a été privilégiée.

Enfin, quatrième exemple, le taux de maintien dans l'emploi des travailleurs handicapés à 6 mois qui est censé mesurer l'activité de l'Association pour la gestion du fonds pour l'insertion professionnelle des personnes handicapées (AGEFIPH). Or mesurer l'atteinte de l'objectif « l'insertion professionnelle des personnes handicapées » par le biais de ce seul indicateur risque de conduire à une sélection des personnes handicapées selon leur probabilité de réussite dans une entreprise.

Les indicateurs inopérants

Parmi les indicateurs inopérants, on peut citer le taux d'annulation par le Conseil d'État des arrêts des Cours administratives d'appel censé constituer un contrôle de qualité. Or, le directeur du programme en question est également le vice-président du Conseil d'État. Si le taux dérive, il pourrait ainsi réunir les présidents de chambre pour leur demander d'être moins sévères. Par ailleurs, la valeur cible du taux d'annulation s'élève à moins de 15 % du total des arrêts des cours administratives, soit une cible largement supérieure aux taux réels de 10,9 % en 2004 et de 11,5 % en 2003.

Les indicateurs indiquant autre chose que leur objet déclaré

Parmi les indicateurs indiquant autre chose que leur objet déclaré, on peut citer celui de la sécurité civile mesurant la disponibilité opérationnelle des canadiens durant la saison des feux. Or cette disponibilité ne résulte pas d'une action de l'État en général et des services de sécurité

civile en particulier, mais de l'action de la société privée X ou Y à qui l'État a sous-traité l'entretien des canadiens.

De la même façon, le nombre de crimes et délits constatés ne mesure pas la criminalité mais l'activité de la police.

Les indicateurs non maîtrisables

Enfin, on peut mentionner les indicateurs non maîtrisables. Il s'agit des indicateurs sur lesquels l'État a peu d'influence, comme le nombre de titres d'information politique et générale, qui correspondent globalement aux publications sérieuses « *apport[ant] de façon permanente sur l'actualité politique et générale, locale, nationale ou internationale, des informations et des commentaires tendant à éclairer le jugement des citoyens* ». Il vaut même mieux que l'État n'ait pas trop d'influence à cet égard dans une démocratie. Ainsi, si une aide publique à la presse existe, cette aide reste générale.

Les indicateurs incompréhensibles aux non-spécialistes

Par ailleurs, certains indicateurs sont incompréhensibles aux non-spécialistes. D'autres indicateurs sont composites. Il est donc difficile d'en saisir et d'en contrôler le contenu.

Les six indicateurs « survivants »

Sur les 32 indicateurs, j'en ai trouvé 6 « survivants », c'est-à-dire qui n'ont pas de défaut flagrant.

Le premier mesure l'indice de satisfaction des usagers du service historique de la défense. Techniquement, cet indicateur ne peut pas être contesté. Toutefois, il est possible de douter de son utilité majeure pour la Défense nationale.

Deuxièmement, l'indicateur du rejet de dioxines par les incinérateurs semble correct. En effet, un plan européen de mise en conformité des incinérateurs a été édicté et il est apparemment respecté.

Troisièmement, j'ai conservé au bénéfice du doute l'indicateur du nombre d'incidents d'exécution des opérations de dette et de trésorerie entre la direction du Trésor et la Banque de France.

Quatrièmement, le délai moyen de paiement des dépenses publiques constitue un indicateur acceptable, bien qu'il soit parfois plus intéressant de surveiller les délais extrêmes qui placent les entreprises en faillite.

Cinquièmement, j'ai supposé que le ministère de la Recherche pouvait influencer l'efficacité de la politique de valorisation de la recherche, qui est calculée par la part des Français parmi les personnes qui déposent des brevets. J'ai donc conservé l'indicateur.

Enfin, le rapport entre le revenu fiscal moyen par unité de consommation en zone urbaine sensible (ZUS) et celui qui est constaté dans les unités urbaines correspondantes mesure le niveau de vie dans les quartiers d'habitat social, ce qui correspond bien à l'objectif de la politique de la ville.

Des indicateurs absents

Mon échantillon n'a pas pris en compte les objectifs transversaux aux programmes.

Par ailleurs, aucun indicateur n'a été prévu pour mesurer le surpeuplement dans les prisons, alors même que la Cour européenne des droits de l'homme condamnait la France à cet égard en 2006, première année d'application de la loi.

Toutefois, il ne faut pas désespérer tout de suite.

Pourquoi les indicateurs LOLF ne sont-ils pas satisfaisants ?

Premièrement, les mêmes indicateurs doivent répondre à trop d'exigences à la fois (encadrés 10 et 11). Au titre de la loi, ils sont censés mesurer les résultats de l'action de l'État. Pour le Parlement, les indicateurs devaient éclairer les choix budgétaires et infléchir la loi de finances. Dans l'Administration, ils devaient servir à juger l'action des responsables administratifs. À cet égard, les primes des directeurs de l'administration centrale et des directeurs des programmes devaient être indexées sur l'atteinte des objectifs et donc sur les indicateurs. Enfin l'administration souhaitait que ces indicateurs permettent de corriger les trajectoires, donc de piloter l'administration.

Deuxièmement, les options d'application de la loi sont contestables. Pourquoi a-t-il été décidé que la loi s'appuierait sur des indicateurs peu nombreux ? Une confusion existe ainsi entre les priorités et les objectifs. En Grande-Bretagne, un dispositif similaire distingue d'une part entre cinquante et cent indicateurs prioritaires qui font l'objet du discours politique, et d'autre part les objectifs de l'action de l'État qui sont considérablement plus nombreux.

ENCADRÉ 10

Ce qui est demandé aux indicateurs LOLF

Des indicateurs aux objectifs multiples

Au titre de la loi : mesurer les résultats de l'action de l'État

Pour le Parlement : éclairer les choix budgétaires

Pour la hiérarchie : juger l'action des responsables administratifs

Pour l'administration : corriger les trajectoires

ENCADRÉ 11

Les options d'application de la LOLF

Des indicateurs qui confondent priorités et objectifs

Des indicateurs voulus peu nombreux

Ne peuvent le plus souvent qu'illustrer des éléments saillants,

Et pas mesurer les résultats (art. 51)?

La confusion entre priorité (peu nombreuses par définition) et objectifs (forcément nombreux)

Quelques propositions pour améliorer l'efficacité des indicateurs LOLF

Tout d'abord, il faudrait fonder le choix des indicateurs sur des résultats d'évaluation des politiques publiques (encadré 12). Tant que la chaîne de résultats n'est pas établie depuis l'action du fonctionnaire de base jusqu'au résultat que la société attend de l'action de l'État et de la puissance publique, il est impossible d'attribuer des indicateurs à chacun des objectifs. Il apparaît en outre nécessaire de distinguer clairement les indicateurs relatifs aux grandes priorités nationales (100 indicateurs) de ceux relatifs aux objectifs de performance de l'action publique (10000 indicateurs). Par exemple, en termes de sécurité routière, l'objectif

visant à réduire le nombre de morts sur les routes ne peut pas être imposé tel quel aux gendarmes. Il leur sera plutôt demandé d'effectuer un certain nombre d'alcotests et de contrôles de vitesse dans des conditions particulières, selon la théorie d'action fondée sur des études affirmant que l'alcool et la vitesse sont des facteurs de risque sur la route et que la peur du gendarme rend sage. Toutefois, il est vrai que les théories d'action pour les affaires étrangères, les affaires sociales, la culture ou la sécurité publique sont plus floues et entraînent légitimement des controverses entre prévention et répression, court terme et long terme, et plus généralement entre les intérêts des uns et ceux des autres, etc.

Ensuite, il faudrait associer les services déconcentrés et les opérateurs au choix des indicateurs .

Enfin, il est nécessaire de débattre démocratiquement de l'action de l'État. Or ces indicateurs sont décidés entre les directeurs des administrations centrales en dehors du regard de toutes les parties prenantes et sans véritable débat. Il serait possible et souhaitable de rendre compte publiquement de l'action de l'État à tous les échelons pertinents de l'administration. Par exemple, les préfets pourraient organiser une conférence de presse annuelle pour faire savoir si les objectifs ont été atteints.

ENCADRÉ 12

Quelques propositions

Fonder le choix des indicateurs sur des résultats d'évaluation des politiques publiques

Associer les services déconcentrés et les opérateurs au choix des indicateurs

Distinguer les grandes priorités nationales (100 indicateurs ?), et les objectifs de performance de l'action publique (10 000 indicateurs ?)

Décliner les « chaînes de résultats » de la base au sommet

Où en est-on quatre ans après la première année d'application de la LOLF ?

Les indicateurs qui ont disparu

15 indicateurs sur 32 n'existent plus. Quatre d'entre eux ont disparu car l'objectif qu'ils reflétaient a été atteint. Quatre autres ont disparu sans explication. Un indicateur a été supprimé pour des raisons techniques. Enfin, six derniers indicateurs ont disparu alors que les performances se dégradent, ce qui alimente le soupçon selon lequel les indicateurs qui reflètent des mauvaises performances sont supprimés en priorité quand il faut réduire le nombre total d'indicateurs.

Les objectifs atteints

Le niveau de certification selon les critères de l'Union européenne (UE) d'un poste de commandement interallié et interarmées a été atteint à 100%. Un plan de mise en conformité pour 2007 avait été planifié au sujet du rejet de dioxines par les incinérateurs. Il a été achevé en 2007, du moins administrativement. S'agissant du nombre d'affaires pénales traitées par magistrat à l'instruction en Tribunal de grande instance, l'objectif de 61 affaires, qui correspondait déjà à la réalité, a été atteint. Enfin, l'indicateur du pourcentage de centres régionaux d'éducation physique et sportive (CREPS) pour lesquels le taux de ressources propres est supérieur ou égal à 50% a été atteint. Il est bon de noter que l'atteinte de cet objectif reflète en fait un désengagement de l'État.

Les disparitions d'objectifs avant que ne soit possible leur comparaison

Chaque année, un programme d'actions prioritaires détaillant les objectifs et les indicateurs correspondants est élaboré. Au printemps de l'année suivante, le rapport annuel de performance (RAP) présente l'évolution de ces objectifs et indicateurs. Les cibles sont souvent définies pour cinq ans, jusqu'en 2011. Or plusieurs objectifs et indicateurs disparaissent d'un document à l'autre, avant qu'il ne soit possible de les comparer à leur cible.

Ainsi, l'objectif et l'indicateur de l'accroissement annuel de la production d'énergie renouvelable thermique, conséquence à la fois des aides de l'agence de l'environnement et de la maîtrise de l'énergie (ADEME) et du crédit d'impôt pour les particuliers, ont disparu lors du renouvellement de la convention de l'ADEME avec l'État.

Par ailleurs, l'indicateur de la pertinence des actions coordonnées d'assistance technique en matière de lutte contre les drogues dans les zones géographiques prioritaires a disparu entre 2007 et 2008. L'objectif visant à favoriser l'adoption d'enfants « à particularités », à savoir les enfants qui présentent un handicap (ces adoptions devant être en nombre et en proportion proche de celles des enfants sans particularités) a également disparu. Il en va de même de l'objectif et l'indicateur de la part des études économiques sur l'industrie française réalisées par l'Institut national de la statistique et des études économiques (INSEE) consacrées à des problématiques prioritaires.

Au final, seule la disparition de l'indicateur du rapport entre le revenu fiscal moyen par unité de consommation en ZUS et celui constaté dans les unités urbaines entre 2006 et 2007 ont été expliqués par le fait qu'il est « *difficile d'établir un lien direct entre son évolution et les performances du programme* ».

Les indicateurs de mauvaises nouvelles

Enfin, 6 indicateurs ont été supprimés parce qu'ils n'évoluaient pas dans le bon sens. Premièrement, l'indicateur du nombre de consultations des sites internet du Département et des postes diplomatiques du ministère des Affaires étrangères a été supprimé car le nombre des visites n'augmentait pas. Il avait même baissé une année.

L'indicateur de la quantité d'azote maîtrisable prise en charge dans le cadre du programme de maîtrise des pollutions d'origine agricole n'était pas non plus satisfaisant, ni celui correspondant à la part des opérateurs du programme dans les brevets déposés à l'Institut national de la propriété industrielle (INPI) par des déposants français.

L'objectif pourtant prudent de rétablissement rapide des collectivités déstabilisées par des circonstances exceptionnelles et le délai de trois ans de réalisation des opérations engagées ont été supprimés car le délai réel augmentait de trois mois chaque année.

Par ailleurs, le taux de maintien dans l'emploi des travailleurs handicapés a été supprimé car il n'atteignait pas l'objectif. En outre, les applicateurs de la loi ont convenu qu'il s'agissait d'un indicateur concernant l'Association nationale pour la gestion du fonds d'insertion professionnelle des handicapés (AGEFIPH) plutôt que l'État.

Enfin, l'objectif et l'indicateur synthétique de connaissance de leurs effectifs par les ministères ont été supprimés avant même qu'un chiffre n'ait été publié. Il est problématique de réduire le nombre des indicateurs au prix d'un repli des objectifs.

Les indicateurs « survivants »

Onze indicateurs se comportent correctement par rapport à la cible. Quatre d'entre eux sont des indicateurs de qualité d'un service rendu. Trois indicateurs stagnent mais doivent leur maintien au positionnement prudent de leur cible. Trois autres ne mesurent pas le résultat

de l'action de l'État. Enfin, un dernier mesure la progression d'une option de politique publique.

Les indicateurs de qualité d'un service rendu

L'indice de satisfaction des usagers du service historique de la défense résiste et progresse.

La progression de l'indicateur d'accessibilité physique des collections au public mesure les efforts fournis par les musées.

L'indicateur de délai de paiement des dépenses publiques semble correct pour le moment.

Enfin, le taux de réclamations contentieuses relatives à la taxe d'habitation traitées dans le délai d'un mois correspond à la cible.

Les cibles « prudentes »

Le taux d'annulation par le Conseil d'État des arrêts des cours administratives d'appel fixé à 15 % était prudent par rapport aux taux réels, bien qu'un taux de 14,9 % ait été atteint en 2009.

Par ailleurs, le ratio entre le nombre de personnes saines et sauvées et le nombre de personnes impliquées après une opération de sauvetage coordonnée par les centres régionaux opérationnels de surveillance et de sauvetage (CROSS) constitue un indicateur relativement hypocrite. En effet, 200 à 250 personnes que les secours ne sont pas parvenus à sauver périssent chaque année en mer. Or il a été décidé de mesurer le ratio entre le nombre de personnes sauvées et le nombre de sorties en mer plutôt que de fixer comme objectif une diminution du nombre de morts. Pour augmenter le ratio, le nombre de sorties en mer a simplement été réduit, alors que le nombre de morts restait identique. En outre, la cible du ratio s'élevait à 97 % pour une réalité de 98,4 %. Autrement dit, alors que 1,6 % de personnes meurent en mer, la cible visait à ne pas dépasser 3 % de morts.

Enfin, l'objectif de stagnation du pourcentage de jeunes conducteurs embauchés grâce aux centres de formation des apprentis (CFA) par rapport aux recrutements totaux est atteint car le pourcentage reste à 31 %.

Les indicateurs « Chantecler » sur lesquels l'État n'a pas d'influence

Le nombre de titres d'information politique et générale croît d'année en année. En outre, s'agissant de la disponibilité opérationnelle de la flotte de la base d'avions de la sécurité civile par flottille durant la « saison feux », le sous-traitant exerce correctement son métier. Enfin le nombre de crimes et délits constatés en zone Police et qui mesure l'activité de la police et non son effet, baisse de 2,5 % par an, conformément à la norme programmée.

Un indicateur d'« option de politique »

L'indicateur du montant d'aide au développement apportée par l'Agence française de développement (AFD) sous forme de prêt par euro de subvention de l'État fonctionne correctement.

Les indicateurs d'échec

Les indicateurs démontrant que la cible n'est pas atteinte comprennent le taux d'établissements recevant du public (ERP) soumis à obligation de contrôle visités par la commission de sécurité. Chaque année, le préfet établit une liste des ERP à visiter. Or seuls 70 % des ERP sont effectivement visités sur un objectif de 100 %.

Par ailleurs, le nombre d'incidents d'exécution des opérations de dette et de trésorerie augmente.

Ensuite, la part des crédits de formation consacrée à la spécialisation des personnels d'adaptation et d'intégration scolaires (AIS) diminue alors que l'objectif visait à l'augmenter.

Le taux de croissance de l'emploi salarié dans les secteurs exonérés de cotisations sociales au titre d'une année rapporté au taux de croissance de l'emploi salarié outre-mer n'est pas satisfaisant.

De même, le pourcentage d'équivalents temps plein (ETP) consacrés aux interventions dans les communes répondant aux critères de l'Assistance technique de l'État au titre de la solidarité et de l'aménagement du territoire (ATESAT) par rapport au total d'ETP de l'ingénierie publique ne progresse pas suffisamment.

Enfin, seuls 70% des plans sanitaires ont donné lieu à des retours d'expérience ou à un exercice au cours des trois dernières années, contre un objectif de 100%.

L'ensemble de ces objectifs et indicateurs de la LOLF ne fait toutefois l'objet d'aucune analyse qui conduirait à une amélioration éventuelle de certains d'entre eux.

À quoi la LOLF devait-elle servir ?

La LOLF devait tout d'abord servir à évaluer les résultats des politiques publiques pour nourrir le débat démocratique et à en tirer des enseignements pour les orientations des politiques futures. Les textes prévoyaient en outre que la LOLF devait permettre de mesurer la « performance » de l'action de l'administration de l'État, afin de récompenser ou de sanctionner les responsables d'une part, et de piloter l'action et infléchir sa conduite au vu des résultats d'autre part.

Les conditions d'un débat démocratique

Pour favoriser le débat démocratique, il aurait fallu présenter une analyse sur le moyen terme de l'évolution des indicateurs. Pourtant, les responsables de programme ne semblent pas vouloir se soumettre au débat sur leurs résultats réels.

Des garanties d'impartialité auraient également été nécessaires, afin d'empêcher que des indicateurs dérangeants puissent être supprimés. Enfin, il aurait été souhaitable que les parties prenantes aux politiques publiques soient un minimum associées à la définition des indicateurs, comme c'est le cas dans les autres pays.

Orienter les politiques

Pour orienter les politiques, il était possible de se servir d'indicateurs d'alerte (encadré 13). Toutefois, ces indicateurs d'alerte sont trop peu nombreux pour infléchir les politiques. À cet égard, il faudrait s'appuyer sur des experts, faire intervenir les parties prenantes et organiser une évaluation pluraliste des politiques publiques.

ENCADRÉ 13

Pour orienter les politiques

Des indicateurs d'alerte peuvent montrer les besoins d'inflexion

Aujourd'hui ils sont :

Peu nombreux et inutilisés

Mais la reconception d'une politique nécessite

Des bases de données lourdes

Des études

Des débats et la formation de consensus

Engager la responsabilité des dirigeants

Afin d'engager la responsabilité des dirigeants, il faut d'abord s'assurer de l'imputabilité de tel objectif à tel dirigeant. Or quand l'imputabilité est réellement définie, peu d'actions relèvent en réalité du ressort de l'État, à l'exception des activités simples mesurées par des indicateurs de service ou des segments simplifiés d'une action complexe comme le respect de normes, de procédures et de plannings.

Piloter l'action et infléchir sa conduite

Le champ d'élection du « système LOLF » ne porte en fait que sur une partie accessoire de l'action de l'État, celle des activités de service pouvant être administrées « dans les conditions du privé ». Cela est dû en grande partie au fait que le management par objectifs impose souplesse et mobilité, ce que la LOLF avait prévu sans que cela ait été suivi dans les faits. Une telle démarche est inapplicable en période de repli des moyens.

Le « système LOLF » ne couvre ainsi pas le cœur de la raison d'être de l'État, à savoir ses activités de souveraineté et de représentation (Affaires étrangères, Défense), ou encore de régulation sociale, d'orientation et d'influence, pour lesquelles l'accessoire est souvent mesurable mais pas l'essentiel.

L'essentiel doit en fait faire l'objet d'une évaluation pluraliste comprenant des études et des débats.

L'approche objectifs/indicateurs pour l'évaluation des politiques sociales

Le cadre européen

Isabelle Engsted-Maquet (Commission européenne)

La méthode ouverte de coordination (MOC) pour l'inclusion sociale et la protection sociale

L'Union européenne n'a aucune compétence directe dans le domaine social. Seuls les États membres possèdent les compétences en la matière. Toutefois, face à leurs défis communs, les chefs d'État et de gouvernement ont estimé qu'il serait utile de définir des objectifs communs non contraignants, laissant chacun libre du choix des moyens pour les atteindre. Ainsi, la méthode ouverte de coordination (MOC) fonctionne autour d'objectifs communs très généraux qui sont formulés de manière consensuelle entre les États membres.

Un ensemble d'outils permet d'évaluer les actions entreprises par les États membres et de constater dans quelle mesure ils parviennent à atteindre les objectifs communs. Ces outils comprennent les indicateurs qui sont liés à des objectifs politiques, mais également des outils de *reporting* par le biais desquels les États membres détaillent les actions qu'ils entreprennent afin d'atteindre les objectifs communs. Ces rapports nationaux stratégiques, que les États membres délivrent tous les trois ans, sont analysés par les services de la Commission européenne dans le cadre d'un rapport conjoint qui est ensuite co-adopté par la Commission et les États membres. Il ne s'agit donc pas d'une évaluation libre et indépendante de la Commission. Enfin, l'outil de financement PROGRESS soutient et promeut la coopération politique et l'apprentissage mutuel. PROGRESS finance ainsi des études, des revues paritaires et d'autres activités qui permettent aux administrations nationales et aux parties prenantes d'échanger dans le cadre de la MOC.

La Commission est favorable à une méthode intégrée, comprenant des objectifs transversaux. Ainsi en 2001, l'inclusion sociale a fait l'objet d'une MOC. Cette méthode a ensuite été élargie à l'enjeu des retraites en 2002, puis à celui de la santé et des soins de longue durée en 2004. En 2005-2006, la Commission a décidé d'intégrer le cadre de la méthode ouverte de coordination pour l'inclusion sociale et la protection sociale.

Le développement des indicateurs : définitions, sources et concepts

L'élaboration des indicateurs

Les indicateurs communs ont été élaborés de manière plus consultative que les indicateurs de la LOLF. Ils ont été développés par des experts représentant les gouvernements des États membres au sein du sous-groupe « Indicateurs » du Comité de la protection sociale (ISG). Ces experts, qui font le lien entre la technique et le politique, ont été appuyés par les services de la Commission, à savoir Eurostat et la direction générale Emploi et Affaires sociales. Ils ont utilisé une méthode collective se basant sur l'observation empirique des données existantes et sur l'expertise au niveau européen. À cet égard, les rapports Atkinson, Marlier, Cantillon et Nolan (2005 et 2007) font référence. D'autres rapports ont été élaborés en 2006 et 2007, tandis que les rapports d'Eurostat et des chercheurs invités ont également été pris en compte.

Le sous-groupe « Indicateurs » a ensuite présenté son travail au Comité de la protection sociale. La liste originale des indicateurs a été adoptée par le Conseil européen de Laeken en 2001, ce qui illustre l'effort politique qui a été accompli autour de ces indicateurs.

L'utilisation des indicateurs

Ces indicateurs sont mis à la disposition de toutes les parties prenantes¹⁰ qui peuvent s'en emparer. Ils sont utilisés dans les rapports nationaux et le rapport conjoint afin, en théorie, d'assurer le suivi des progrès réalisés et éventuellement d'établir des cibles. En réalité, de nombreux indicateurs n'étaient pas disponibles pour mesurer les tendances. Les indicateurs sont également utilisés pour comparer les pays. Enfin, le sous-groupe « Indicateurs » a essayé de développer un cadre analytique pour établir un diagnostic et identifier les domaines d'action prioritaires, comme la pauvreté des enfants qui sera détaillée dans l'exemple de *benchmarking*.

Les critères de sélection des indicateurs

Si l'on en croit les critères de sélection qui ont été établis en début d'exercice, le portfolio des indicateurs doit être complet et couvrir toutes les dimensions de manière équilibrée. Il doit également être intégré, c'est-à-dire prendre en compte les liens entre les dimensions. Enfin, il doit être transparent et compréhensible.

Les indicateurs eux-mêmes doivent être pertinents, mesurables et donner lieu à une interprétation claire et unanimement reconnue. Ils doivent en outre être robustes et validés statistiquement. Les statistiques qui les sous-tendent doivent être comparables entre les pays et dans le temps. Les indicateurs doivent par ailleurs reposer sur des données disponibles rapidement. Enfin, ils doivent être sensibles aux effets des politiques sans qu'il soit possible de les manipuler et sans qu'ils n'entraînent d'effets pervers.

Une typologie d'indicateurs

En réalité, nous ne trouvons pas toujours un indicateur à la fois disponible, comparable, etc. qui serait nécessaire pour décrire une dimension donnée. Le sous-groupe « Indicateurs » a donc défini des modalités permettant d'atteindre les nécessaires compromis. Ainsi, la typo-

10. Sur le site internet : http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_and_social_policy_indicators/introduction

logie d'indicateurs labellisés « EU » qui a été adoptée couvre quasiment tous les critères. Ces indicateurs permettent de comparer les États membres en niveau et en tendance et reflètent leur situation réelle. Il s'agit d'indicateurs de résultats sociaux (*social outcomes, intermediate social outcomes* ou *outputs*).

Par ailleurs, les indicateurs qui ne répondent pas à tous les critères sont classés sous le label « NAT ». Bien qu'ils soient basés sur des définitions communes et agréées et reflètent des dimensions clés, ils ne permettent pas de comparaison directe entre les États membres ou ne donnent pas lieu à une interprétation claire et normative.

Par exemple, il est important de pouvoir mesurer l'évolution des dépenses de sécurité sociale. Toutefois, l'Union européenne n'a pas encore fixé le pourcentage idéal de dépenses de sécurité sociale. De même, de nombreux éléments de contexte sont nécessaires pour interpréter un indicateur tentant de prédire l'évolution du taux de remplacement théorique des retraites par rapport au revenu d'activité. Aucune cible unique ne peut être ainsi définie de manière consensuelle dans le domaine des dépenses de sécurité sociale. Pourtant, ces indicateurs « NAT » sont très importants pour mesurer la nature et la taille de l'intervention publique. Ils ne peuvent être interprétés qu'en connaissance d'un contexte associé. D'autres indicateurs « NAT », tels que l'espérance de vie en bonne santé, sont difficiles à mesurer du fait de biais culturels entre les États qui acceptent d'être comparés en tendance mais pas en niveau.

Pour compléter les indicateurs « NAT », nous disposons également d'informations de contexte.

Enfin, les indicateurs nationaux permettent d'étudier des dimensions telles que les minorités ethniques ou les migrants, pour lesquelles aucune donnée harmonisée au niveau européen n'existe. Il est alors demandé aux États membres de rendre compte de ces situations sur la base d'indicateurs nationaux, bien qu'ils ne soient pas comparables.

Définitions et sources

Nous nous référons si possible aux définitions internationales. Par exemple, pour étudier les taux de pauvreté, nous utilisons une définition du revenu qui a été établie par le groupe de Canberra en 2001. De la même façon, pour mesurer l'emploi, nous retenons la définition du Bureau International du Travail (BIT). Toutefois, des sujets restent en suspens, parmi lesquels figurent le traitement des loyers imputés et l'évaluation des biens autoproduits.

Parmi les sources dont nous disposons, le système statistique européen comprend l'enquête EU-SILC sur les revenus et les conditions de vie, la *Labour Force Survey* (LFS) relative à la force de travail et l'enquête ESPROSS concernant les dépenses de protection sociale. Quand le système statistique européen ne peut pas être utilisé, nous utilisons les sources des autres organisations internationales bien que plusieurs de ces sources soient l'objet de débats entre certains États membres. Parmi ces sources on peut citer les enquêtes de l'OCDE, PISA pour l'éducation ou SHA pour la santé.

L'enquête EU-SILC constitue la base de référence pour la moitié de nos indicateurs. Elle a l'avantage d'être légalement contraignante. Toutefois, l'harmonisation est réalisée *ex post*. Les critères de qualité compris dans EU-SILC permettent d'obtenir des données relativement fiables et difficilement contestables par les États membres. EU-SILC comprend une dimension panel par rotation qui n'a toutefois pas encore été explorée. En outre, des modules existent sur des thèmes spécifiques. EU-SILC constitue donc une grande avancée, bien que des limites apparaissent. La méthode ouverte de coordination est un utilisateur privilégié de cette source et peut influencer son développement. Ainsi, le sous-groupe « Indicateurs » pourra contribuer à la révision prochaine d'EU-SILC.

Exemple de concept commun : la pauvreté

Au-delà d'un consensus apparent au niveau européen, la manière de mesurer la pauvreté est très débattue entre les États membres. En 1975, le Conseil des ministres s'est accordé sur une définition selon laquelle les pauvres correspondent aux individus ou aux familles dont les ressources sont si limitées qu'elles les empêchent de bénéficier du niveau de vie minimum acceptable dans leur pays. Selon cette définition qui est très précieuse pour les statisticiens, la pauvreté est relative au niveau de vie de l'État membre dans lequel l'individu vit. Ainsi, ce principe justifie les taux de pauvreté relatifs avec un seuil national, qui sont toutefois très débattus entre les États membres.

Suite aux multiples travaux sur les indicateurs de pauvreté selon la définition de 1975, l'Union européenne est parvenue à s'accorder sur une nouvelle définition de la pauvreté relative au pays de résidence, pondérée selon l'intensité de l'écart entre le revenu médian des pauvres et le revenu médian national, multidimensionnelle et temporelle, en référence notamment à la pauvreté persistante sur plusieurs années.

L'indicateur de pauvreté est par ailleurs sensible à un ensemble de politiques macro-économiques censées être utiles pour lutter contre cette pauvreté. Par exemple, le taux de pauvreté monétaire est sensible aux politiques de redistribution et d'emploi.

En revanche, ce concept de pauvreté a été remis en question avec l'accession de nouveaux États membres qui a pour conséquence l'existence de grandes différences de niveau de vie en Europe.

Les indicateurs

Les objectifs transversaux

Les deux premiers objectifs transversaux sont très vastes. Ils couvrent quasiment tous les champs d'action de la politique sociale. Premièrement, l'objectif de cohésion sociale, qui n'est pas encore précisément défini, sera mesuré par des indicateurs en termes d'inégalité de revenu, de pauvreté, de distribution territoriale de l'emploi, etc. Afin d'atteindre cette cohésion sociale, il faut mettre en place un système de protection sociale adéquat, accessible, efficace et durable financièrement et socialement. Le deuxième objectif transversal exige que ces politiques de protection et d'intégration sociale interagissent mutuellement et positivement avec les politiques de l'emploi et de croissance.

Pour ces deux premiers objectifs, quatorze indicateurs de résultats ont été retenus. Ils évaluent :

- le taux et l'intensité de la pauvreté ;
- les inégalités de revenu ;
- l'espérance de vie en bonne santé ;
- le nombre de sorties précoces du système scolaire ;
- le taux de ménages sans emploi ;
- la projection des dépenses de sécurité sociale ;
- l'adéquation des retraites ;
- l'accès aux soins, en mesurant le pourcentage de personnes affirmant ne pas avoir consulté un médecin bien qu'elles en eussent besoin, en raison d'un coût trop élevé, de listes d'attentes trop longues ou d'une trop grande distance d'accès ;
- le taux de pauvreté ancré dans le temps, relatif à l'effet de la croissance sur le niveau de vie ;
- le taux d'activité général ;

- le taux d'emploi des personnes âgées ;
- la pauvreté des personnes participant au marché du travail. En effet, plusieurs pays possèdent des systèmes de protection sociale qui ont augmenté le nombre de personnes dépendant de prestations d'invalidité au cours de ces dernières années ;
- la dispersion des taux d'emploi régionaux en l'absence de taux de pauvreté régionaux ;
- les dépenses de santé publique.

En outre, douze indicateurs de contexte aident à comprendre ces quatorze indicateurs de résultats (encadré 14).

ENCADRÉ 14

Les douze indicateurs de contexte

- | | |
|--|--|
| (1) GDP growth | (8) Jobless households by main household types |
| (2) Employment rate,
Unemployment rate,
Long term unemployment share | (9) Making work pay indicators unemployment trap, inactivity trap, low-wage trap |
| (3) Life expectancy at 0, 65 | (10) Net income of social assistance recipients |
| (4) Old age dependency ratio, current and projected | (11) At-risk of poverty rate before social transfers (other than pensions) |
| (5) Distribution of population by household types | (12) Change in projected theoretical replacement ratio + change in projected public pension expenditure 2004-2050. |
| (6) Public debt, current and projected | |
| (7) Social protection expenditure, current, by function, gross and net | |

Les principaux résultats

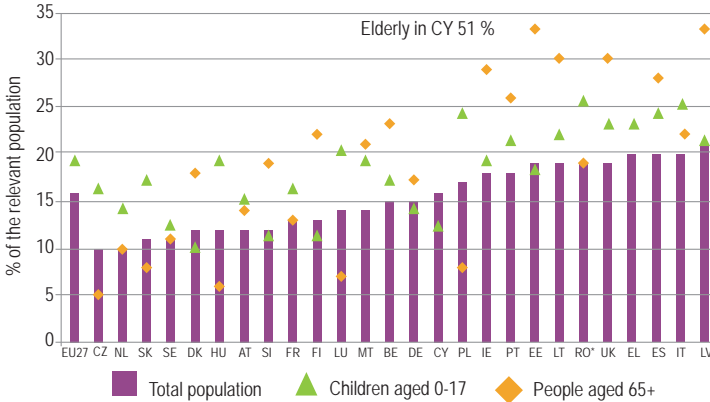
À titre d'exemple, nous présentons ici les principaux résultats portant sur le taux de pauvreté, l'impact des transferts sociaux, le revenu net des bénéficiaires de l'assurance sociale et l'accès aux soins.

En 2007 (année de collecte), le taux de pauvreté le plus bas a été enregistré en République tchèque tandis que les taux de pauvreté les plus élevés concernent l'Italie et la Lituanie (graphique 12). Ainsi, si cet indicateur permet d'évaluer pour chaque pays la manière dont les revenus sont redistribués envers les plus démunis, il ne permet toutefois pas de comparer les pays.

Par ailleurs, l'impact des transferts sociaux, à l'exception des pensions de retraite, sur la réduction de pauvreté est mesuré par le biais de deux indicateurs (graphique 13). Ainsi, le taux de pauvreté après les transferts sociaux en 2007 est d'environ 17% moins élevé qu'avant en Espagne, en Grèce et en Italie, selon EU-SILC 2007. En revanche, l'impact des transferts sociaux atteint presque 60% dans les pays nordiques et en Hongrie. Cet indicateur, bien que contestable car il mesure des taux de pauvreté théoriques, est utile pour lancer le débat au sujet de la différence des moyens qui sont consacrés par les États membres à la réduction de la pauvreté.

GRAPHIQUE 12

Taux de pauvreté

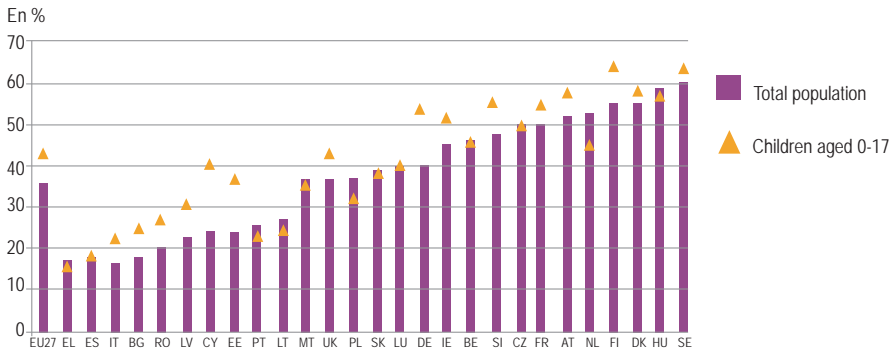


Lecture • Le taux de pauvreté s'établissait à 14 % au Luxembourg pour la population générale, 20% pour les mineurs et 6,5% pour les personnes de plus de 65 ans.

Source • EU-SILC 2007 ; income year 2006 ; except for UK (income year 2007) and for IE (moving income reference period 2006-2007) ; BG : National household budget survey 2006 ; RO : National household budget survey 2007.

GRAPHIQUE 13

Impact des transferts sociaux : réduction du risque de pauvreté en pourcentage du taux de pauvreté avant transferts



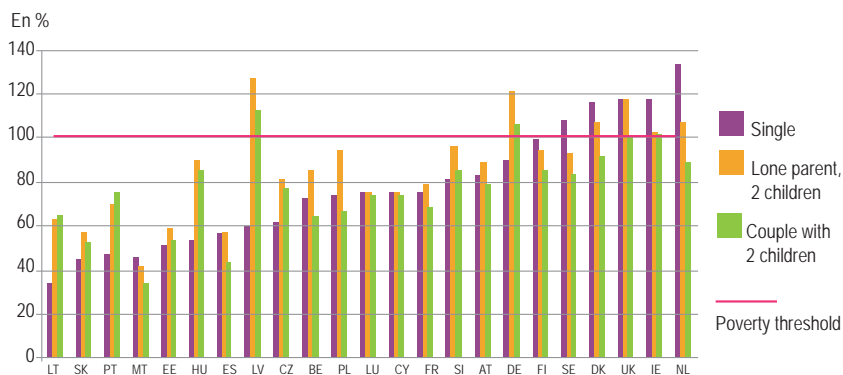
Lecture • Le taux de pauvreté après transferts sociaux (hors pensions) est 40 % moins élevé (52 % en ce qui concerne les mineurs) qu'avant en Allemagne.

Source • SILC 2007 ; income year 2006 (income year 2007 for IE and the UK) ; BG : National household budget survey 2006 ; RO : National household budget survey 2007.

Ensuite, l'indicateur du revenu net des bénéficiaires de l'assistance comparé au seuil de pauvreté permet de constater les écarts de traitement entre les pays selon qu'il s'agit d'individus seuls, de parents isolés avec enfants ou de couples avec enfants (graphique 14).

GRAPHIQUE 14

Revenu net des bénéficiaires de l'assistance en pourcentage du seuil de pauvreté



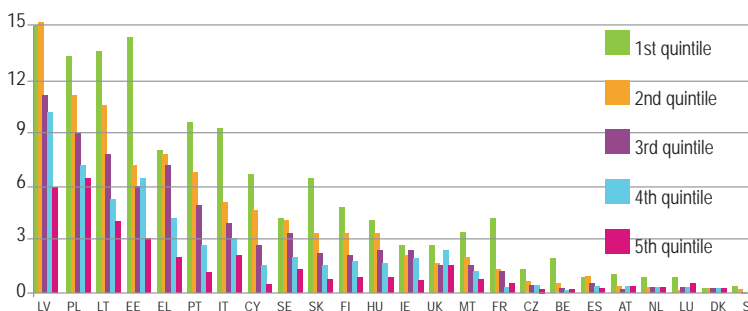
Lecture • En Pologne, le revenu net des bénéficiaires de l'assistance s'établissait à 72% du seuil de pauvreté pour les personnes célibataires, 95% pour un parent isolé avec 2 enfants et 67% pour un couple avec deux enfants.

Source • SILC 2007 ; income year 2006 (income year 2007 for IE and the UK) ; BG : National household budget survey 2006 ; RO : National household budget survey 2007-OCDE/EC tax benefit model 2007.

L'indicateur d'accès aux soins par quintile de revenus (graphique 15) permet enfin de constater l'écart très important qui existe entre le premier et le dernier quintile, sauf au Royaume-Uni où le service public de la santé est entièrement gratuit.

GRAPHIQUE 15

Indicateur d'accès aux soins



Lecture • Au Portugal, le taux de personnes interrogées ayant déclaré avoir renoncé à des soins s'établissait à 9,5% pour les ménages appartenant au premier quintile de revenus et à 1% pour ceux du 5^e quintile.

Source • SILC 2006 ; income year (income year 2006 for IE and the UK) ; data missing for BG, DE and RO.

Un exemple de *benchmarking* européen sur la pauvreté des enfants

La méthode du *benchmarking* sur la pauvreté des enfants

En analysant en détail les indicateurs et en les mettant en relation avec une analyse plus approfondie de comparaison internationale, les principaux déterminants de la pauvreté des enfants ont pu être identifiés.

Les trois types de facteurs majeurs concernent :

- les caractéristiques sociodémographiques des ménages, notamment leur taille, leur composition (parents isolés, familles nombreuses...), l'âge et le niveau d'éducation des parents ;
- la situation des parents sur le marché du travail (taux d'emploi...) ;
- l'impact de l'action gouvernementale, comprenant notamment l'existence d'aides familiales et la manière dont le système d'impôts et de prestations interagit avec le marché du travail.

Comme les États n'apprécient pas de faire l'objet d'un classement dans le domaine de la politique sociale, notre tableau évalue les pays à la fois selon l'intensité de la pauvreté des enfants qu'ils présentent par rapport à la moyenne européenne et par rapport à la situation du reste de leur population, et selon les trois principaux facteurs de pauvreté. Trois indicateurs clés ont été choisis pour illustrer ces facteurs, à savoir le pourcentage des enfants appartenant à des foyers sans emploi, le pourcentage des enfants dont les parents travaillent mais sont quand même pauvres, et l'impact des transferts sociaux.

Les pays n'ont pas été regroupés en fonction de la performance des États en matière de pauvreté des enfants, mais en fonction de leur profil au regard des trois facteurs (tableau 6).

TABLEAU 6

Performance des États en termes de prévention de la pauvreté des enfants

		Child poverty risk outcomes	Children in jobless households	In-work poverty	Impact of social transfers
Group A	AT	+	+	++	++
	CY	+++	+	+++	+
	DK	+++	+	+++	++
	FI	+++	++	+++	+++
	NL	+	+	+	+
	SE	+	(++)	++	++
	SI	++	+++	+++	++
Group B	BE	+	--	+++	+
	CZ	-	--	+	+
	DE	++	--	+++	+++
	EE	--	--	+	-
	FR	++	-	++	++
	IE	-	---	+	+
	SK	-	---	+	+

		Child poverty risk outcomes	Children in jobsless	In-work poverty	Impact of social transfers
Group C	HU	---	---	-	+
	MT	-	--	--	-
	UK	--	---	--	+
Group D	EL	--	+++	--	---
	ES	---	+	---	---
	IT	---	++	---	--
	LT	---	+	--	--
	LU	--	+++	--	+
	LV	---	-	--	--
	PL	---	-	--	--
	PT	--	+	--	--

De (---) très mauvaise à (+++) très bonne.

Le groupe A comprend les pays où peu d'enfants appartiennent à des foyers sans emploi, où les parents qui travaillent ne sont pas pauvres et où l'impact des transferts sociaux sur la réduction de la pauvreté est important par rapport aux autres pays européens. Quasiment tous les pays de ce premier groupe affichent de bonnes performances en matière de pauvreté des enfants.

Les pays du **groupe B** comprennent de nombreux enfants vivant dans des foyers sans emploi. En revanche, l'impact des transferts sociaux et le taux de pauvreté au travail sont relativement satisfaisants dans ces pays. Le problème principal de ces pays concerne l'accès au marché du travail de certains types de ménages comme les parents isolés. La question qui se pose pour ces pays est celle de la segmentation du marché du travail et d'un éventuel problème de prestations sociales qui décourageraient certains ménages à rester sur le marché de l'emploi.

Le groupe C comprend les pays présentant des taux élevés de ménages sans emploi et de pauvreté au travail, et un faible impact des transferts sociaux sur la pauvreté des enfants. Au Royaume-Uni, de nombreuses familles sont monoparentales. Or, les prestations sont relativement généreuses. En outre, les gardes d'enfants sont rares ou fonctionnent seulement à temps partiel, ce qui empêche ou décourage de nombreuses familles monoparentales d'accéder au marché du travail. Par ailleurs, les foyers monoparentaux concernent des femmes très jeunes par rapport aux autres pays européens.

Le groupe D englobe les pays dont l'impact des transferts sociaux sur la pauvreté des enfants est nul car ces transferts comprennent peu d'allocations familiales. La pauvreté au travail est également très élevée dans ces pays. En revanche, aucun ou très peu d'enfants appartiennent à des ménages sans emploi, souvent parce que les parents n'ont pas d'autre choix que de travailler. Ces pays comprennent très peu de familles monoparentales. Il est probable qu'en l'absence de prestations sociales, une femme élevant seule ses enfants n'a pas d'autre choix que de travailler ou de vivre chez ses parents, d'où l'existence de nombreux ménages complexes et multi-générationnels. Dans ces pays, le problème majeur concerne le soutien aux familles.

L'utilité de l'exercice de benchmarking

Cet exercice d'analyse a permis de lancer le débat en exposant les différences entre les pays et en laissant la responsabilité de l'interprétation de ces résultats aux différents pays membres ainsi qu'aux chercheurs. Sur la base de ce tableau, un chercheur a ainsi analysé la situation en Irlande en étudiant l'impact des différentes politiques sociales sur les indicateurs répertoriés. L'exercice de *benchmarking* peut donc contribuer à l'élaboration d'une théorie de l'action.

L'analyse de la pauvreté des enfants a été publiée dans le rapport conjoint et a permis d'évaluer les nouvelles politiques. Parfois, il a été constaté que les États avaient pris ou envisagé des mesures cohérentes avec les résultats de l'analyse.

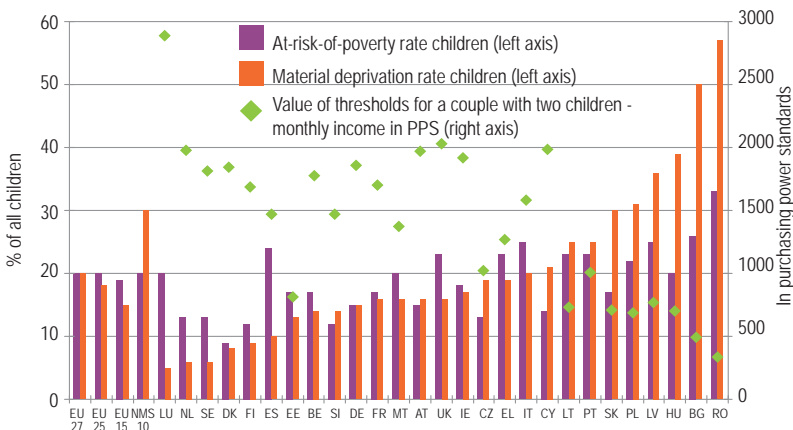
Un cadre analytique commun renforce la légitimité de l'exercice d'évaluation au niveau européen. Par ailleurs, il peut favoriser l'apprentissage mutuel dans la mesure où les pays peuvent évaluer la manière dont les États qui ont le même profil qu'eux ou ceux qui présentent les meilleurs profils se comportent. En termes de communication, ce cadre analytique a permis à la Commission de justifier son message appelant à agir à la fois sur l'accès des parents au marché du travail et sur les prestations sociales. En effet, les États membres qui se portent le mieux sont ceux qui combinent des politiques intelligentes de soutien à l'accès au marché du travail et un minimum de soutien au revenu des familles. Le consensus sur le message est plus facilement obtenu quand il s'agit d'un sujet concernant les enfants.

Certains pays se sont étonnés de leurs performances par rapport à d'autres pays. Le cadre analytique leur permet en effet d'évaluer l'ampleur du problème grâce à une comparaison internationale.

Les indicateurs sélectionnés mesuraient seulement la pauvreté relative. Toutefois, un indicateur de pauvreté des enfants en termes de conditions de vie a également été développé (graphique 16). Cet indicateur composite, qui était utilisé par la France avant d'être utilisé au niveau européen, mesure les conditions de vie par le biais de diverses privations. La classification des pays selon cet indicateur en 2008 place en première position les anciens États membres, qui disposent de PIB par tête élevés par rapport aux nouveaux États membres.

GRAPHIQUE 16

Taux de pauvreté relatif et taux de pauvreté en conditions de vie



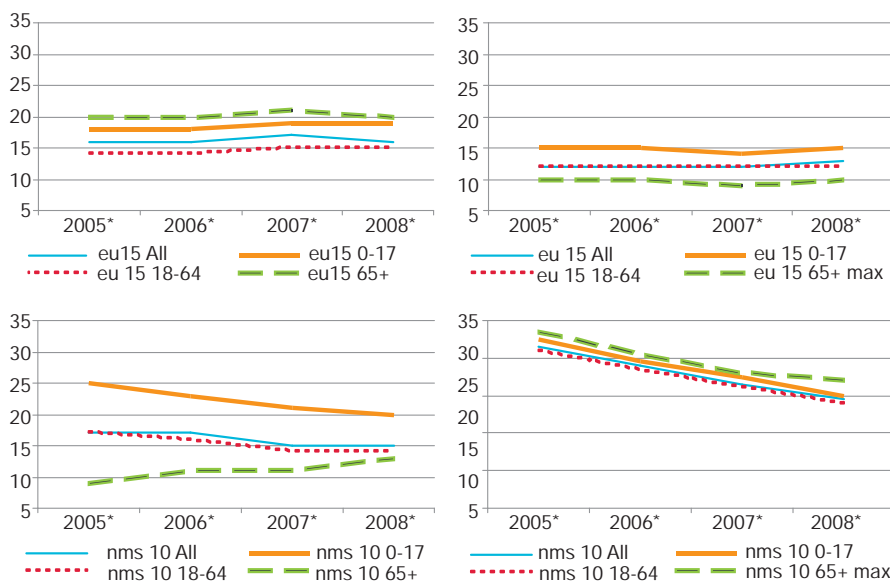
Lecture • La pauvreté en conditions de vie des enfants s'élevait à 15% en Allemagne.

Source • EU-SILC 2008.

En termes d'évolution, le constat est différent. Entre 2005 et 2008 (année de collecte des données), la pauvreté relative et le taux de privations matérielles stagnent au sein de l'Union européenne à quinze (graphiques 17). En revanche dans les nouveaux États membres, la pauvreté en conditions de vie diminue très significativement. En outre, le taux de pauvreté relatif des enfants diminue de 25 % à 20 %, tandis que le taux de pauvreté relatif des personnes âgées augmente de 6 % à 13 %, voire de 20 % à 50 % dans certains pays. Cette tendance particulièrement marquée dans les nouveaux États membres s'explique par une très forte croissance des salaires, tandis que les pensions indexées sur l'inflation ont augmenté dans une moindre mesure. Ainsi, le pouvoir d'achat relatif des personnes âgées a diminué et leur taux de pauvreté a considérablement augmenté sous l'effet combiné de l'augmentation du seuil de pauvreté et de la baisse de leur situation relative. En termes de pauvreté en conditions de vie, cependant, la situation des personnes âgées dans les nouveaux États membres progresse moins vite que celle des autres catégories de la population.

GRAPHIQUE 17

Évolution du taux de pauvreté relatif et de la pauvreté en conditions de vie



* Ces années sont celles de la collecte des données.

Lecture • Anciens pays membres (EU-15) en haut, nouveaux pays membres en bas, taux de pauvreté relatif à gauche et pauvreté en conditions de vie à droite.

Source • EU-SILC 2005, 2006, 2007, 2008.

SYNTHÈSE DES DÉBATS AVEC LA SALLE

Distinction entre indicateurs de résultat et indicateurs de gestion

Question de la salle

Les démarches objectifs/résultats peuvent-elles servir à la fois à mesurer les résultats finaux des politiques publiques sur le niveau et les conditions de vie des citoyens, et à responsabiliser les administrations chargées de la mise en œuvre de ces politiques ?

En effet d'une part, l'expérience montre qu'il est très difficile de responsabiliser nos opérateurs au-delà de l'efficience des services qu'ils produisent. Or bien que l'efficience des services soit importante, elle ne reflète pas la totalité de l'efficacité finale des politiques.

D'autre part, nos démarches objectifs/résultats sont accolées aux lois financières. Or ces lois financières servent rarement de support à l'expression de la responsabilité politique des objectifs finaux de ces démarches. L'exercice du budget vise plutôt l'allocation des ressources.

Peut-être la bonne démarche consisterait-elle à procéder en deux temps. Tout d'abord, il faudrait définir des indicateurs finaux correspondant aux objectifs finaux des politiques publiques et les suivre sur le moyen terme. Ensuite, dans les procédures d'allocation des ressources, il faudrait décliner à l'attention des opérateurs qui mettent en œuvre les politiques publiques, des objectifs opérationnels assortis d'indicateurs mesurant réellement l'impact de l'action de ces opérateurs. Le lien entre ces objectifs opérationnels et les objectifs globaux qui sont assignés aux politiques devrait en outre être explicité.

Par exemple, la loi hôpital, patients, santé et territoire (HPST) stipule que l'État et les caisses d'Assurance maladie doivent négocier et conclure un contrat qui servira ensuite de cadre pour les relations entre les personnels de santé et les organismes de sécurité sociale. Ce contrat devrait constituer le document dans lequel sont fixés les objectifs globaux de la politique de l'Assurance maladie, tandis que des objectifs précis seraient déclinés dans les conventions d'objectifs et de gestion (COG) entre l'État et chacune des caisses. Cette démarche pourrait éventuellement permettre de réconcilier le suivi des objectifs finaux des politiques et le suivi de l'action des opérateurs.

Réponse de M. Brunetière

Cette méthode reflète ce qui doit effectivement être mis en œuvre. Dans le cas de la politique sociale, qui est une politique majeure et complexe, les indicateurs d'état tels que l'indicateur de pauvreté doivent être mesurés de manière indépendante. En parallèle, un débat social est toutefois nécessaire afin de déterminer précisément dans quelle mesure et par quels moyens la pauvreté doit être diminuée. Ce débat donne lieu à une somme de résultats précis qu'il faut atteindre. La politique étant toutefois un lieu de contradictions, certaines mesures ne pourront pas être mises en œuvre car elles s'opposent aux objectifs d'autres politiques. Par exemple, l'augmentation des impôts sur le revenu nécessaire à la redistribution de la richesse envers les pauvres s'opposera aux objectifs contraires d'une autre politique visant à réduire les impôts.

Question complémentaire de la salle

La loi relative à la politique de santé publique visait à définir des résultats politiques sur la base d'une évaluation périodique de la santé de la population par rapport à un ensemble d'indicateurs de santé importants. Or il devient difficile de définir des résultats politiques quand un indicateur LOLF est chargé de juger à la fois l'impact de la politique et la performance d'une partie des acteurs qui contribuent à l'évolution de cet indicateur d'impact. Il faut faire l'effort de construire la théorie d'action et de repérer un indicateur adéquat pour mesurer la responsabilité des acteurs. Je pense que le débat est essentiel. Toutefois, des points de repère sont nécessaires pour ancrer ce débat.

Réponse de M. Brunetière

Les chiffres et le débat sont en effet tous deux importants. Le même indicateur ne peut pas répondre à toutes les questions à la fois. Chaque notion possède un sens qui diffère de celui de la notion voisine et qui nécessite une instrumentalisation différente. Fondamentalement, les indicateurs sont divisés en deux catégories. Les uns observent un état du monde tandis que les autres permettent de suivre l'effet des décisions ou des mesures qui sont prises. Or ces deux catégories d'objectifs obéissent à des logiques différentes, notamment à l'échelle du temps.

Il convient ainsi d'opérer la distinction entre les indicateurs de résultat final et les indicateurs de résultats intermédiaires qui sont liés à une chaîne de résultats. En théorie, les indicateurs de la LOLF devaient mesurer des résultats intermédiaires. L'audit des indicateurs devait donc non seulement juger de l'atteinte des résultats en fonction des objectifs, mais également vérifier que l'administration était dotée du degré d'action correspondant à l'atteinte des résultats prévus. Or les deux niveaux ont été confondus.

Le choix de la LOLF consistait à mesurer les résultats intermédiaires. Toutefois, un phénomène d'évitement s'est produit. Un travail est en ce moment en cours d'exécution au sujet de la liaison entre l'objectif et l'indicateur. Il s'agit de constater si l'indicateur indique vraiment un résultat qui correspond à l'objectif. Or des « chaînons manquants » existent entre certains objectifs et leurs indicateurs. Par exemple, le rayonnement de la France constitue une finalité ou un « super objectif », tandis que l'indicateur qui lui est associé évalue seulement le nombre de visites sur le site internet du ministère. Il s'agit en réalité d'un exemple d'évitement. En effet, les dépenses pour la défense du français à l'étranger sont de moins en moins importantes. Or politiquement, il n'est pas possible d'annoncer une telle réalité.

Par ailleurs, une différence existe entre la branche Maladie et la branche Allocations familiales. En tant que gestionnaires, la branche Maladie fonctionne en grande partie en termes d'indicateurs et d'objectifs, du fait notamment de l'existence de multiples conventions comportant de nombreux chiffres, et en raison de l'objectivation des indicateurs de santé. En effet, les indicateurs de santé permettent d'alerter, puis de discuter d'une politique et enfin de confier à certains acteurs des actions de prévention sur un problème de santé publique, des recherches sur un médicament, etc. La branche Allocations familiales s'appuie quant à elle plutôt sur un ensemble de données permettant d'alimenter le débat public et de préciser la conception des politiques publiques. Notons que ces données permettent d'établir des comparaisons internationales et de produire du *benchmarking*. À cet égard, les indicateurs Programme de qualité et d'efficience (PQE) fonctionnent incontestablement mieux que ceux de la LOLF.

Complément de la salle

La différence principale entre la branche Maladie et les autres branches concerne la problématique de la gestion des risques. En effet, les caisses d'Assurance maladie exercent un rôle en termes de conditionnement et d'action sur les comportements des professionnels de santé et des patients. En revanche dans le cadre des autres branches, comme celle des retraites notamment, la responsabilité de la gestion des risques est imputée au législateur. Le recouplement entre les PQE et les objectifs des Conventions d'objectifs et de gestion (COG) paraît donc trop faible dans ce cas.

Nombre des indicateurs

Complément de Mme Brocas, directrice de la DREES

Peut-être que la volonté de définir peu d'indicateurs constitue une fausse bonne idée. Dans l'analyse qui a été menée au niveau européen, l'indicateur de pauvreté ancrée dans le temps se rapproche d'un indicateur de pauvreté absolue. Or d'un côté, plusieurs États membres critiquent l'approche consistant à évaluer l'évolution dans le temps de la pauvreté dans un pays donné. De l'autre, la comparaison internationale des taux de pauvreté relative, qui démontre par exemple que la République tchèque présente un taux de pauvreté relative plus faible que celui de la France, est également critiquée et conduit certains à penser qu'il faut plutôt étudier la pauvreté absolue.

Pour comprendre une politique, il faut tout d'abord identifier le problème qui doit être réglé socialement et l'action publique qui doit être mise en œuvre afin d'y remédier. Or il est possible de commencer à identifier des problématiques relatives à un problème donné grâce à la juxtaposition d'une multitude d'indicateurs qui nécessiteront ensuite des études plus approfondies et la mise en œuvre d'actions publiques.

Réaction de M. Brunetière

Dans les rapports de la commission Stiglitz, le PIB constituait à l'origine l'unique indicateur. Or à mesure que les travaux avançaient, les membres de la commission Stiglitz se sont rendus compte que plusieurs chiffres étaient nécessaires pour décrire un monde de plus en plus complexe. Toutefois, un chiffre sans sa métadonnée associée ne vaut rien, de même que l'exposition des chiffres sans le débat nécessaire à leur sujet n'est pas non plus valable. Les chiffres n'existent que pour permettre, par le biais d'une discussion démocratique, de se repérer dans un monde complexe.

La politique représente le lieu des volontés contraires, et donc du double langage. S'agissant de l'Assurance maladie par exemple, les remboursements ne semblent pas augmenter en proportion, ce qui n'empêche pas les discours politiques de plaider pour une amélioration. Le double langage politique perdurera toujours.

Théorie de l'action

Question de Mme Brocas, directrice de la DREES

D'après vous, il faut avoir défini une théorie d'action pour ensuite pouvoir construire les objectifs et les indicateurs. Or est-il toujours possible de définir une théorie de l'action ?

En effet, certaines politiques font l'objet d'un compromis et, en conséquence, comprennent plusieurs objectifs qui sont parfois contradictoires. Devons-nous chercher une rationalité opératoire pour l'élaboration des indicateurs et des objectifs, ou plutôt accepter que les politiques publiques englobent intrinsèquement plusieurs objectifs, voire résultent d'un compromis entre des acteurs qui ne défendent pas automatiquement la même théorie d'action ?

Réponse de M. Brunetière

La théorie d'action peut manquer de clarté dans plusieurs politiques en raison de la difficulté du sujet ou d'une concurrence entre plusieurs théories d'action. Par exemple, dans le cas de la politique visant à éviter la récidive des malfaiteurs, une théorie d'action suggère de les maintenir enfermés tandis qu'une autre propose de leur apprendre à ne pas récidiver quand ils sont en prison, et qu'une autre encore affirme qu'il aurait fallu en faire de bons citoyens avant qu'ils ne se retrouvent en prison. En démocratie, le sujet est débattu et donne lieu à des décisions.

Par ailleurs, l'existence d'options contradictoires paraît normale dans le cadre d'une démocratie qui favorise le débat, la confrontation, le rapport de forces et les jeux d'acteurs. Gérer les contradictions constitue le cœur de l'action de la puissance publique. Sinon, nous n'aurions pas besoin d'un État à qui nous confions le pouvoir d'exercer la violence légitime.

Question de la salle

Il est important de distinguer l'évaluation de l'action publique de l'évaluation de la gestion publique. En effet, l'action publique relève de la responsabilité du gouvernement devant le Parlement. En revanche, la gestion publique concerne l'administration qui s'engage devant le gouvernement. Si les indicateurs politiques macro et les objectifs finaux peuvent faire l'objet d'un débat, les indicateurs pour responsabiliser les acteurs ne doivent pas nécessairement donner lieu à un tel débat.

Par ailleurs, certaines théories d'action sont définies au niveau politique, comme le choix entre prévenir et réprimer en matière de sécurité. Pour responsabiliser les acteurs, il faut tenir compte de la capacité de gestion des acteurs car l'imputabilité des résultats n'est pas automatique.

Réponse de M. Brunetière

Je ne pense pas que le débat au sujet du management, et donc de l'exécution des politiques, soit inutile. Durant l'évaluation de la politique de la sécurité routière en 2000-2001, certains procureurs affirmaient être impitoyables en matière de délits d'alcool mais laxistes envers les délits de vitesse. Chaque fonctionnaire agit en fonction de ses propres représentations du monde et de sa compréhension des consignes qui lui sont données. Or l'action publique dépend en réalité du travail des fonctionnaires. Dans le domaine du management durable, nous avons réalisé un audit dans tous nos services au sujet de la compréhension du concept par les agents. Nous avons découvert un univers mental qui devra évoluer parce qu'il constitue la réalité de l'action publique, des jugements, des autorisations, des arbitrages, des influences, des normes et des ressentis concernant l'administration, des rapports entre les personnes, etc. Nous devons gérer de moins en moins de robots et de plus en plus d'imaginaires. Nous sommes donc obligés de débattre de l'ac-

tion publique à l'intérieur des administrations. L'indicateur sert à illustrer et à produire de l'imaginaire au sein des administrations.

Contractualisation entre parlementaires et administration

Question de la salle

Ces difficultés ne résultent-elles pas d'un manque de contractualisation entre les parlementaires et l'administration sur la fonction même des indicateurs et sur la définition des concepts ?

Réponse de M. Brunetière

Les États évitent parfois de définir précisément certains concepts car ils n'apprécient pas les classements pour tous les indicateurs.

La LOLF a permis de déplacer le clivage entre le législatif et l'exécutif, qui n'est plus très pertinent en raison du fait majoritaire, vers un clivage entre l'administration et le politique dans lequel les instances politiques, comprenant le gouvernement et le Parlement, ont conclu un contrat avec l'appareil administratif d'État. Le Parlement souhaitait en effet que les responsables des programmes, à savoir les directeurs des administrations centrales, rendent compte de leurs « performances » directement devant le Parlement, en l'absence du cabinet. Toutefois, pour éviter que les parties n'évitent le processus, il faut qu'une autre partie en surveille le bon déroulement.

Pilotage financier autorisé par la LOLF

Question de Mme Brocas, directrice de la DREES

La difficulté que la LOLF a rencontrée en termes d'identification des objectifs résulte du fait qu'elle était censée fixer à la fois des objectifs à la gestion publique et des objectifs de politique publique. La démarche de la LOLF comprend toutefois deux innovations très importantes. D'une part, il a été décidé que le budget serait voté au premier euro. Deuxièmement, la LOLF a estimé que la gestion par programmes constituerait une démarche efficiente permettant notamment de redéployer, dans un cadre financier contraint, les moyens d'un programme vers un autre. Ce système favorise-t-il en réalité le redéploiement de moyens d'un programme vers un autre ?

Réponse de M. Brunetière

La question des services votés ou du premier euro m'a toujours paru plutôt théorique et formelle car certaines dépenses sont inéluctables.

Par ailleurs, la notion de programme est relativement dévoyée. À l'origine, l'idée consistait à allouer pendant une période donnée une allocation à une catégorie particulière afin de suivre l'évolution de la catégorie donnée. Juste avant la fin du programme, les décideurs politiques se réunissaient pour décider si un autre programme était nécessaire. La notion de programme est limitée dans le temps. Peu de pays placent l'intégralité de leur budget dans ce système de programmes. En revanche, la France a voulu lancer des programmes dans plusieurs domaines au même moment qu'elle décidait de la refonte des allocations

budgétaires et du management par objectifs et indicateurs sur l'ensemble de l'administration, y compris sur les secteurs qui ne s'y prêtent pas. En effet, les programmes sont adaptés dans certains secteurs mais pas dans d'autres. Par ailleurs, dans le secteur du développement durable, le programme énergie-climat qui vise à lutter contre le réchauffement climatique dépend également de tous les autres programmes. En effet, un résultat de politique publique ne dépend pas d'un seul secteur de l'administration. Le rayonnement de la France à l'étranger dépend de tous les Français et de quelques étrangers, des artistes, des industriels, des touristes, etc. Ces acteurs concourent également à d'autres objectifs comme l'équilibre de la balance commerciale. L'imputabilité n'existe pas dans le cadre des politiques réservées à l'État. Elle existe quand l'État mène des actions accessoires.

Question de la salle

Dans quatre ans quand cette étude sera menée à nouveau, vous rencontrerez les mêmes problèmes concernant les indicateurs et les objectifs.

Réponse de M. Brunetière

Il ne s'agit pas d'une fatalité. Par exemple dans le cadre du Grenelle de l'environnement, nous avons rassemblé quasiment toutes les personnes concernées par le sujet du développement durable. En termes de production d'objectifs de politique publique, les discussions qui ont été menées durant une année ont été efficaces. Des feuilles de route comprenant des objectifs ont été définies. Il sera intéressant de constater dans quatre ans ce qu'est devenu le système de gestion des suites du Grenelle de l'environnement. Une représentation de la société civile a préparé le travail et métabolisé les problèmes à destination des parlementaires qui, ensuite, ont voté les lois de Grenelle I, II, etc. Enfin, l'administration et toutes les autres parties prenantes de l'action publique, comme les organisations non gouvernementales (ONG), se sont mises en mouvement de façon cohérente afin d'agir avec efficacité.

BIBLIOGRAPHIE

- Atkinson T., Marlier E., Cantillon B, Nolan B., 2005, *Taking forward the EU Social Inclusion Process*, An Independent Report Commissioned by the Luxembourg Presidency of the Council of the European Union.**
- Atkinson T., Marlier E., Cantillon B, Nolan B., 2007, *The EU and social inclusion : facing the challenge*, The Policy Press, University of Bristol.**
- Brunetière J.-R., 2006, « Les indicateurs de la loi organique relative aux lois de finances (LOLF) : une occasion de débat démocratique ? », *Revue française d'administration publique*, n° 117 p. 95 à 111.**
- Cantillon B., Deleeck H, Van den Bosch K., 2001, « Pour une norme européenne de la pauvreté. Recommandations », *Revue belge de sécurité sociale*, n° 2.**
- Elbaum M., 2009, « Les indicateurs de performance en matière de politiques sociales : Quel sens pour l'action publique ? », *Revue de l'OFCE*, n° 111 p. 81 à 101.**

L'évaluation s'est imposée comme un domaine à part entière des sciences sociales, recouvrant notamment la recherche évaluative, qui vise à identifier les effets propres d'une politique, en les distinguant des effets de contexte ou produits par d'autres politiques menées concomitamment. Au cours de l'année 2009-2010, un séminaire sur les méthodes de l'évaluation a été organisé par la DREES, chaque séance présentant des outils de l'économie et de la sociologie pouvant contribuer à l'évaluation.

La première séance s'est attachée à définir précisément l'évaluation. Les trois séances suivantes ont abordé successivement : les méthodes économétriques visant à identifier les effets propres d'une politique ; les changements organisationnels, tels qu'ils sont perçus à travers les études de terrain ; les méthodes qui permettent d'impliquer les usagers, par l'intermédiaire de focus groups et d'enquêtes de satisfaction. La dernière séance s'est intéressée aux approches consistant à associer objectifs, indicateurs et benchmarking.

L'objet de ce recueil est de retranscrire le plus fidèlement possible les différents exposés de ce séminaire ainsi que les débats qui les ont suivis.



N° DICOM : 11-025