
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.

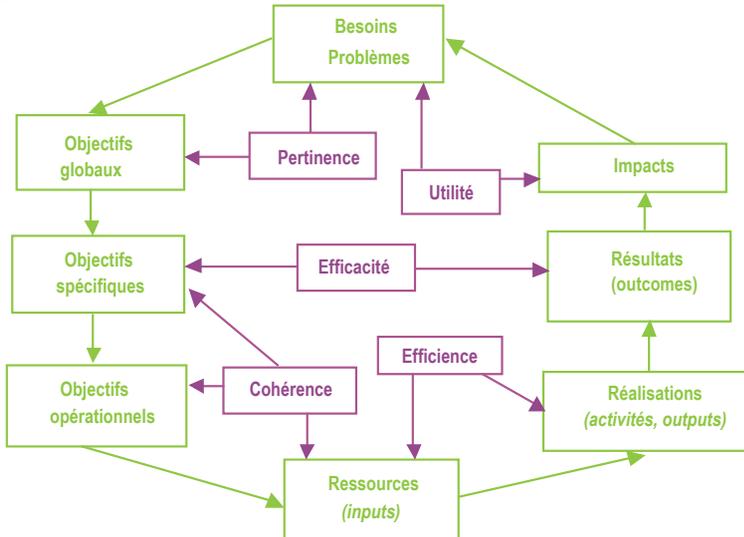
L'évaluation de l'action collective peut porter tout autant sur un projet que sur un dispositif de politique publique, un programme ou des politiques plus larges. Plus le degré de généralité est élevé, plus l'évaluation devient complexe. Pour autant, il est indispensable de procéder à ces évaluations 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 Dufló (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

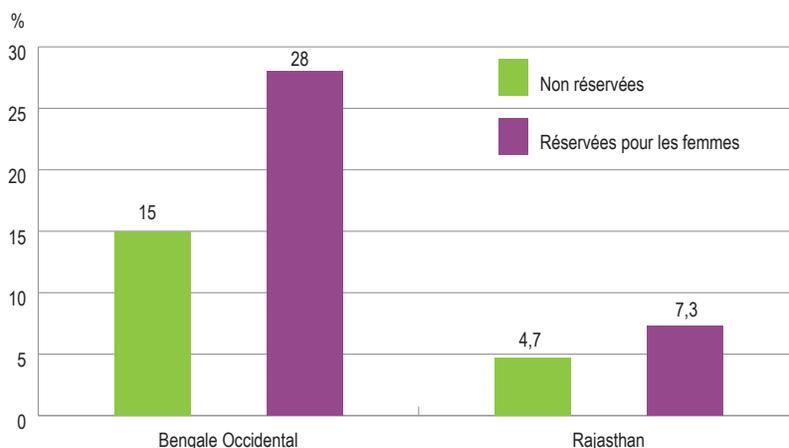
La deuxième possibilité réside dans les expériences naturelles, qui exploitent des situations dans lesquelles les groupes sont parfaitement comparables. Pour reprendre l'exemple de la parité, il est possible de comparer une commune où une femme a gagné de justesse à une autre où une femme a perdu de 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.

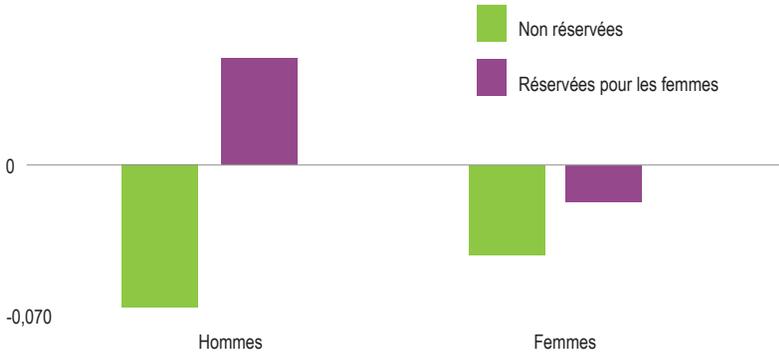
Dans le cas de l'évaluation de la politique de parité hommes-femmes en Inde, la « randomisation » a été mise en place pour des raisons politiques et non pour faciliter le travail. En effet, en 1993, un amendement à la constitution prévoit que les conseils municipaux et les conseils de villages devront 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



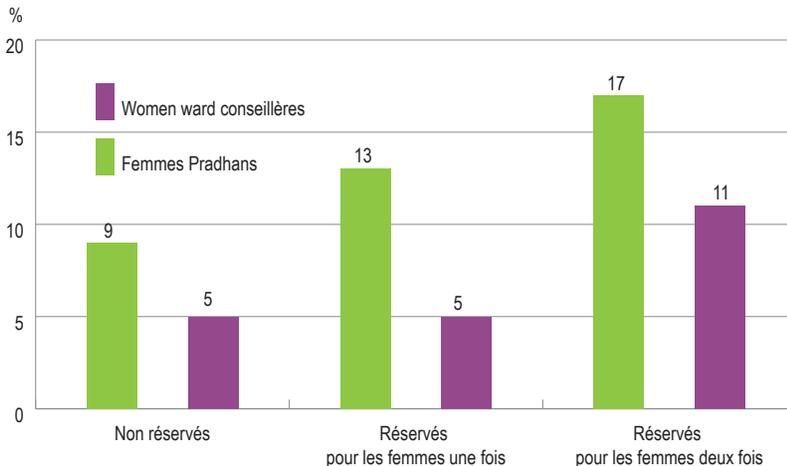
Source • Beaman, Chattopadhyay, Duflo, Pande et Topalova, 2008.

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.

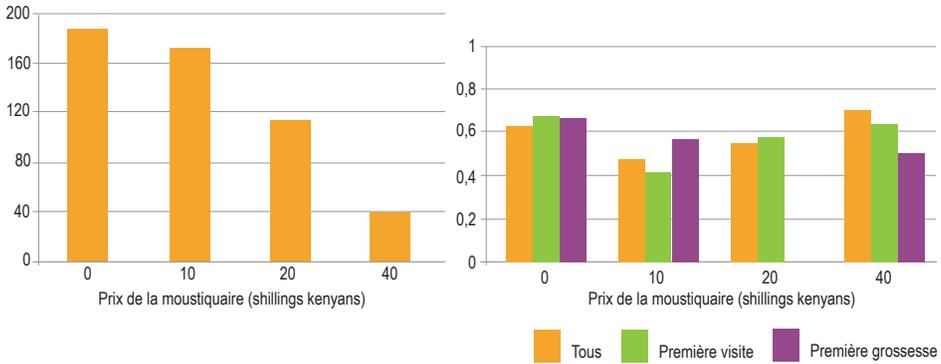
La gratuité comme solution au paludisme

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.

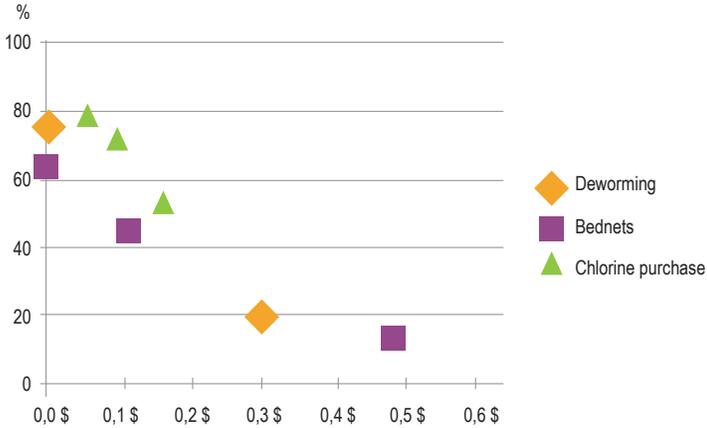
Dani Rodrik publia alors les résultats de cette étude sur un blog très lu sous le titre « Jeff Sachs a raison », Jeff Sachs étant un économiste proche de l'OMS. Sur ce même blog, un autre économiste, Mead Over, protesta, avançant l'argument que ces résultats ne valaient que pour les femmes enceintes 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)



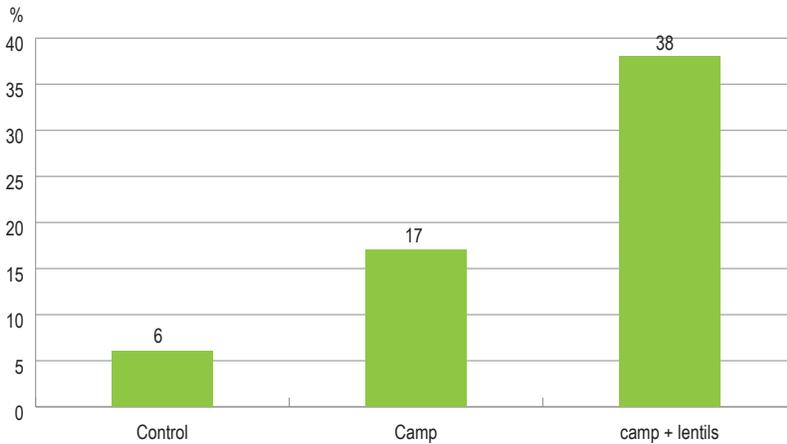
Lecture • 70 % des personnes achètent un traitement à base de chlore coûtant 0,10 \$ (Deworming = déparasitage, Bednets = moustiquair, Chlorine purchase = Traitement à base de chlore).

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.

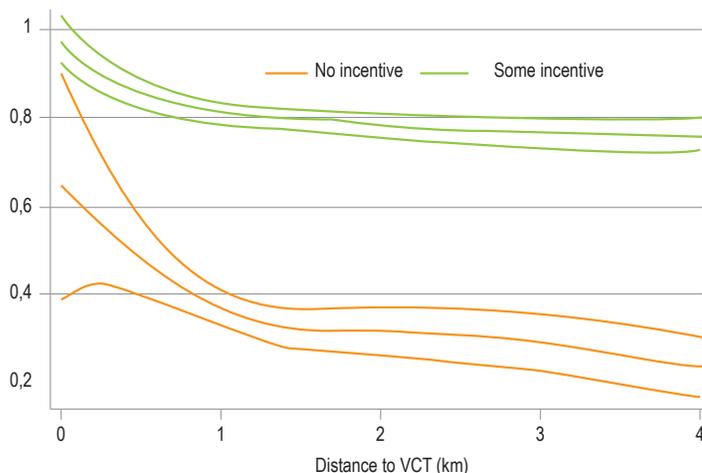
Lecture • Dans les villages concernés par la campagne de vaccination mais où aucune incitation n'a été mise en place, 17 % des enfants ont reçu au moins deux vaccinations.

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-