

L'évaluation des politiques publiques : État(s) de l'art et controverses

Annie Fouquet

DANS **REVUE FRANÇAISE D'ADMINISTRATION PUBLIQUE** 2013/4 N° 148 , PAGES 835 À 847
ÉDITIONS **INSTITUT NATIONAL DU SERVICE PUBLIC**

ISSN 0152-7401

DOI 10.3917/rfap.148.0835

Date de mise en ligne : 25/03/2014

Article disponible en ligne à l'adresse

<https://droit.cairn.info/revue-francaise-d-administration-publique-2013-4-page-835?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Institut national du service public.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

L'ÉVALUATION DES POLITIQUES PUBLIQUES : ÉTAT(S) DE L'ART ET CONTROVERSES

Annie FOUQUET

*Présidente d'honneur de la Société française de l'évaluation**

Résumé

Après une apparente éclipse, l'évaluation des politiques publiques est revenue dans le discours public depuis 2007. La réforme constitutionnelle de 2008 qui inscrit l'évaluation des politiques publiques dans le texte fondamental de la République, les initiatives gouvernementales successives, le renouveau dans les méthodes d'évaluation des impacts traduisent-ils un véritable renouveau de l'évaluation en France ? Ce dossier thématique s'appuie sur les travaux des Journées françaises de l'évaluation, qui se sont tenues en juin 2013 à Grenoble. Leur programme visait à faire le point sur les diverses manières dont l'évaluation se pratique et innove, et affronter les controverses dont elle fait l'objet. Alternant des articles de fond sur l'évaluation en France, des témoignages sur les pratiques, et des points de vue réflexifs, ce dossier décrit dans une première partie les enjeux actuels de l'évaluation en France ; une deuxième partie porte des regards critiques, notamment sur la question de la mesure, et du lien délicat entre évaluation individuelle et évaluation collective. Enfin une troisième partie ouvre le champ des critères d'évaluation comme autant de pistes de recherche.

Mots-clefs

Évaluation des politiques publiques, modernisation de l'action publique, performance, mesure

Abstract

— *Introduction – Since 2007 – and seemingly after a brief lull – public policy evaluation is increasingly present in public discourse. This is evidenced by the constitutional reform of 2008 that made provisions for the evaluation of public policies within the very text which acts as the backbone of the French Republic, the series of governmental initiatives enhancing the role of evaluation in public policy and the renewal of impact-assessment methods: have these actually translated into a resurgence for the domain of evaluation for France? This thematic compilation relies on the work carried out during the Journées Françaises de l'Evaluation, which*

* Annie Fouquet, administrateur de l'INSEE, inspectrice générale honoraire des Affaires sociales, présidente du conseil scientifique de l'évaluation du Fonds social européen, du conseil scientifique de l'Agence nationale d'évaluation des établissements sociaux et médico-sociaux, et du conseil scientifique de l'ONZUS (observatoire des zones urbaines sensibles), après avoir été notamment directrice de la DARES (direction de l'Animation de la recherche, des études et des statistiques au ministère de l'Emploi), du Centre d'études de l'emploi, de la division « études sociales » à l'INSEE.

were held in Grenoble during June 2013. Their programme aimed to take stock of the different ways in which evaluation is performed and innovates. It also addressed the controversies surrounding this topic. By using substantive articles on current evaluation trends in France, testimonials on leading practices as well as a reflexive approach, this article addresses two main points: firstly, the current issues raised within the field of evaluation in France, and secondly, the critical examination of key issues such as measurement, and the delicate relation between individual evaluation and collective evaluation. Finally, it will explore evaluation criteria and suggest avenues for future research.

Keywords

Public policy evaluation, public action modernisation, performance, measurement

« Évaluation des politiques publiques : états de l'art et controverses », tel était le thème choisi par les organisateurs des XI^e Journées françaises de l'évaluation qui se sont tenues à Grenoble en juin 2013¹. L'objectif du programme était de prendre du recul par rapport aux pratiques existantes et faire un retour réflexif et théorique sur l'état de l'évaluation en France, à un moment où l'évaluation des politiques publiques revient en force dans le débat public sous les influences convergentes de décisions gouvernementales depuis 2007 et de méthodes médiatisées d'évaluation des impacts².

Au cours de ces dernières années, tandis que certains se réjouissaient de la volonté de réformer le pays en s'appuyant sur une évaluation faisant la part à l'analyse de la complexité des situations concrètes, dans le même temps, cette insistance gouvernementale, laissant peu de place aux acteurs, était perçue par d'autres comme donnant une image confuse de l'évaluation des politiques publiques, en l'assimilant et la réduisant parfois à du contrôle de gestion, voire à de simples suivis quantitatifs d'indicateurs, ou au contrôle-sanction. Cela a engendré une image négative de l'évaluation dans certains milieux, notamment chez les universitaires, les psychologues, ou les pédagogues : « l'évaluation tue » a-t-on pu écrire, mêlant évaluation individuelle assimilée à la sanction et évaluation d'action publique.

Or, faut-il le rappeler, l'évaluation de politiques, de programmes, ou de pans entiers de l'action publique, si elle peut utiliser les divers outils d'analyse de la dépense publique (indicateurs, audits, contrôle de gestion), elle ne s'y réduit pas et sa finalité est plus large. L'évaluation vise à produire des connaissances en vue, pour les citoyens, d'apprécier la valeur d'une politique (d'un programme, d'un dispositif), notamment quant à ses effets, et, pour les décideurs, de les aider à en améliorer sa pertinence, son efficacité, son efficience, sa cohérence et ses impacts (Charte de la société française de l'évaluation d'après les décrets de 1990 et 1998).

Ce dossier thématique décrit dans une première partie les enjeux actuels de l'évaluation en France ; une deuxième partie porte des regards critiques, notamment sur la question de la mesure et du lien délicat entre évaluation individuelle et évaluation collective. Enfin une troisième partie ouvre le champ des critères d'évaluation comme autant de pistes de recherche

1. Le présent dossier comprend à la fois un certain nombre de communications présentées à l'occasion des 9^e journées de la Société française de l'évaluation, ainsi que des articles soumis directement à la *Revue*.

2. La question des études d'impact fera l'objet d'un dossier thématique dans le prochain numéro de la *Revue française d'administration publique*.

LE RENOUVEAU DANS L'ESPACE PUBLIC

La réforme constitutionnelle de 2008 a donné une place institutionnelle forte à l'évaluation des politiques publiques. En l'inscrivant dans le texte fondamental de la République, en attribuant un rôle au Parlement et à la Cour des comptes, elle met délibérément l'évaluation des politiques publiques dans la sphère législative et dans celle de la plus haute autorité indépendante. Ce faisant, la réforme sort l'évaluation du seul bon vouloir de l'exécutif, dont elle était restée tributaire dans les expériences précédentes, que ce soit celle des années 1990 avec le dispositif placé auprès du Premier ministre (Comité interministériel de l'évaluation/Conseil scientifique de l'évaluation puis Conseil national de l'évaluation), retracé en détail dans leur article par Clément Lacouette et Pierre Lascoumes, ou que ce soit plus récemment en 2007 par la création d'un éphémère Secrétariat d'État à l'évaluation et à la prospective, ou encore dans l'actualité la plus récente, avec le programme d'évaluations lancé par le Premier Ministre dans le cadre de la modernisation de l'action publique (MAP).

Ce dossier s'ouvre par un témoignage de Didier Migaud, Premier Président de la Cour des comptes, qui fut l'un des pères fondateurs de la LOLF³ lorsqu'il était président de la Commission des finances de l'Assemblée. Dans son discours d'ouverture aux Journées, il rappelle les trois finalités de l'évaluation : éclairer la décision, rendre compte au citoyen (article 15 de la Constitution), alimenter la réflexion et le débat public. Aujourd'hui, plusieurs facteurs concourent à rendre l'évaluation nécessaire et utile : le souci de transparence, l'exigence accrue des citoyens quant à la qualité des services publics, et la complexité croissante des politiques publiques, le tout dans un contexte de fortes contraintes sur les finances publiques. Mais le développement de l'évaluation en France nécessite également un changement culturel qu'il appelle de ses vœux. Pour cela, cinq défis majeurs sont à relever : la question des normes et de la définition de l'évaluation ; le positionnement des évaluateurs, notamment leur indépendance ; le niveau et la diversité de leur expertise ; la question des moyens et du temps nécessaire ; et, enfin, le lien entre évaluation, appropriation par les acteurs de la politique et prise de décision politique.

Sur ce dernier point, le lien avec la décision, la Révision générale des politiques publiques (RGPP), lancée par le Président de la République en 2007 (concernant les dépenses de l'État), puis la Modernisation de l'action publique (MAP), lancée par le gouvernement suivant en 2012, ont pris acte du fait que l'évaluation doit rester au plus près du décideur pour être suivie d'effet. La RGPP était pilotée par le Secrétaire général de l'Élysée, ce qui était de bon augure pour la prise de décision mais le protocole d'évaluation laissait volontairement de côté les acteurs, à commencer par les ministres en charge de ces politiques. Ainsi, alors qu'au départ la volonté était de rendre publiques⁴ les conclusions des évaluateurs (équipes mixtes, consultants et inspecteurs généraux), des « erreurs de communication »⁵ ont vite eu raison de ce processus. Au terme de quelques mois, le premier volet de la RGPP, celui qui concernait les politiques, a été abandonné aux ministères qui l'ont enterré et, seul, le second volet, celui de la fonction de production de

3. Loi organique relative aux lois de finances du 1^{er} août 2001 n°2001-692

4. Par la mise en ligne des présentations visuelles des évaluations et de leurs préconisations qui fut brutalement interrompue.

5. Une proposition technique (transférer le financement de la carte « familles nombreuses » de la SNCF à la Caisse d'allocations familiales) a été comprise comme une décision politique portant atteinte à la politique familiale.

l'État, a été mis en œuvre : l'objectif de non-remplacement d'un fonctionnaire sur deux a pris la place de l'analyse des missions, déséquilibrant tout l'édifice.

En octobre 2012, le gouvernement de Jean-Marc Ayrault reprend la volonté de s'appuyer sur l'évaluation des politiques publiques pour moderniser l'action publique. La MAP prend acte des erreurs de la RGPP⁶ : son champ concerne désormais la totalité des dépenses publiques (État, collectivités locales, protection sociale) ; elle responsabilise les ministres, tout en conservant la décision au niveau du Premier Ministre qui préside le Comité interministériel pour la modernisation de l'action publique (CIMAP)⁷. L'ambition de la MAP est de passer en revue, au terme de cinq années, la totalité des dépenses publiques de l'État, des collectivités locales ou associations (protection sociale), soit 1 120 milliards d'euros. Une première vague a engagé 49 évaluations représentant 250 milliards d'euros, soit plus de 20 % de la dépense publique.

*Les évaluations conduites dans le cadre de la Modernisation
de l'action publique, selon le SGMAP⁹*

Elles ont pour finalité de « refonder les politiques publiques en partant des attentes des usagers et en tenant compte des ressources », soit plus précisément : refonder les objectifs de ces politiques, pour s'assurer qu'ils correspondent aux besoins prioritaires ; optimiser les dispositifs existants, afin de garantir que chaque acteur est impliqué à bon escient ; dégager, chaque fois que possible, les économies rendues nécessaires par l'objectif de redressement des finances publiques

Elles doivent produire « des scénarios de transformation résultant d'une démarche partagée ». Pour cela elles devront : établir de manière indépendante un diagnostic sur la pertinence, l'utilité et l'efficacité ; partager ce diagnostic avec l'ensemble des acteurs concernés ; documenter des scénarios de transformation pour éclairer la décision ; définir des plans d'action permettant la mise en œuvre immédiate (dès 2014)

Il s'agit d'introduire durablement dans le processus de décision publique le questionnement évaluatif. Pour cela, chaque ministre est responsable personnellement du choix des sujets d'évaluation, de la mise en œuvre des travaux et de la présentation du scénario retenu en Comité interministériel pour la modernisation de l'action publique. Il devra associer les parties prenantes et mettre en place d'un comité de pilotage pour chaque évaluation.

Pour homogénéiser les pratiques et accompagner les ministères, un guide a été rédigé par une équipe interministérielle d'inspecteurs généraux ; reprenant ce qui se fait en matière d'évaluation au plan international et l'expérience accumulée, ce guide contribue à diffuser une culture évaluative (IGA, IGF, IGAS, déc. 2012). Le secrétariat de ces évaluations est porté par la Secrétariat général à la modernisation de l'action publique (SGMAP) directement rattaché au Premier ministre⁹, dans un partage des

6. Cf. Rapport IGA, IGF, IGAS d'octobre 2012

7. Décret n° 2012-1199 du 30 octobre 2012 portant création du comité interministériel pour la modernisation de l'action publique.

8. Source : Intervention d'Alban de Loisy, chef du département Evaluation des politiques publiques au Secrétariat général à la modernisation de l'action publique (SGMAP), lors des XI^{èmes} Journées françaises de l'évaluation, Grenoble, juin 2013.

9. Décret n° 2012-1198 du 30 octobre 2012 portant création du secrétariat général pour la modernisation de l'action publique.

tâches négocié avec le Commissariat général à la stratégie et à la prospective (CGSP), qui intervient en amont et en aval des évaluations.

Les différentes réunions du CIMAP (2 avril, 17 juillet et 18 décembre 2013) ont permis de constater le déploiement effectif du programme. Réalisées selon les prescriptions du guide et sous des formes adaptées à chaque cas, une vingtaine des 49 évaluations de la première vague initiées le 2 avril sont achevées à la fin de l'année 2013¹⁰. Une seconde vague lancée le 18 décembre 2013, qui porte sur douze nouvelles évaluations, met l'accent sur la réduction attendue de la dépense publique. Il est trop tôt pour juger de cette nouvelle vague d'évaluations gouvernementales¹¹.

L'ÉVALUATION S'INSTITUTIONNALISE ET SA PRATIQUE SE DIFFUSE

Ces initiatives récentes sont-elles le fruit récurrent du « marronnier de l'action gouvernementale » comme s'interrogent Clément Lacouette Fougère et Pierre Lascoumes ? Ces derniers replacent ces initiatives dans la longue lignée des tentatives successives d'améliorer la gouvernance de l'action publique et décrivent l'histoire de l'évaluation en France sous quatre angles : (1) l'évaluation utilisée comme moyen de légitimation et de contre-pouvoir en partant de l'exemple des États-Unis où l'évaluation a permis de justifier des politiques sociales controversées lors du *New Deal* puis dans les années soixante ; (2) l'évaluation utilisée pour améliorer la mise en œuvre des programmes publics, avec la « Rénovation du service public » ; (3) l'évaluation utilisée pour améliorer les politiques sectorielles avec des exemples pris aux ministères de la culture et de l'équipement ; (4) l'évaluation utilisée comme levier d'optimisation des dépenses publiques avec les exemples de la RCB, et plus récents de la LOLF puis de la RGPP.

Au terme de ce parcours dans les différentes scènes où l'évaluation a été mobilisée, les auteurs reviennent sur leur constat initial d'une institutionnalisation qui serait faible, faute de convergence des acteurs et d'accord sur une définition universelle de l'évaluation, pour convenir que ces différentes expériences ont contribué globalement à diffuser les pratiques et à en assurer la continuité. Ils constatent des avancées importantes sur le plan des méthodes, dans la capacité à différencier l'évaluation des démarches voisines : audit, contrôle, inspection ; dans la diversité reconnue des formes d'évaluation ; dans le recours à de nouveaux acteurs : l'évaluation-expertise est aujourd'hui complétée par des démarches ouvertes aux diverses parties-prenantes. Les débats anciens opposant évaluation dite scientifique, évaluation managériale et évaluation démocratique semblent aujourd'hui dépassés. Un langage et des catégories communes se construisent et se consolident, ainsi que des standards de bonnes pratiques.

Vingt années de pratiques de l'évaluation ont permis de préciser les apports respectifs, souvent complémentaires, des différents regards portés sur l'action publique (SFE 2009, Gibert 2010). Contrôle pur (de type inspection), audits, contrôle de gestion (pilotage) et

10. Sur la politique territoriale de l'emploi, l'aide juridictionnelle, le soutien aux entreprises, la formation professionnelle des demandeurs d'emploi, la politique de l'eau, le soutien au sport professionnel...

11. Même si devant cette apparente reprise en mains de la MAP par Bercy qui dicte par avance le résultat de l'évaluation, on peut légitimement s'interroger sur la validité de l'armistice conclu précédemment entre le SG-MAP et le CGSP. On peut se demander si le Commissariat ne devrait pas reprendre la main et retrouver le rôle qui fut celui du Commissariat général au plan dans les années 1990 en matière d'évaluation.

évaluation se différencient par leur finalité, par le cadre de référence auquel l'action observée est comparée et par leurs conséquences. De façon idéal-typique, on peut distinguer :

— le contrôle qui vérifie la conformité aux règles obligatoires (lois, réglementation, tel le cadre comptable) et entraîne, en cas de non-respect, une sanction (amende, procédure judiciaire) ;

— l'audit dont la finalité est de réduire les risques encourus par une organisation¹² (que ce soit en terme organisationnels, financiers...), en référence aux pratiques de la profession ; les conséquences d'un audit sont des recommandations suggérées au commanditaire ;

— le contrôle de gestion (pilotage)¹³ qui a pour objet de suivre la mise en œuvre et la réalisation des actions programmées ; le référentiel en est alors le programme établi en début de période ; et l'analyse des écarts se conduit dans un dialogue de gestion.

L'évaluation, quant à elle, a pour objectif d'optimiser une politique en analysant les écarts entre, d'une part, les résultats qui en sont attendus pour répondre aux besoins de la société et, d'autre part, les résultats qui ont été obtenus. Le référentiel n'est pas fixé à l'avance (contrairement aux standards d'une profession comme dans l'audit) : il dépend des objectifs poursuivis par l'action publique et des finalités attendues de cette évaluation particulière. Quel *evaluanda*, que cherche-t-on à évaluer, quel était l'objet de la politique ? Attend-on une analyse de la mise en œuvre ? Une analyse des résultats ? De la façon dont ils ont été obtenus ? De leur imputabilité ? Attend-on des scénarios prospectifs ? Autant de questions qu'il importe de se poser.

L'évaluation s'appuie, quand ils existent, sur les indicateurs de suivi (pilotage) qui donnent les « réalisations » du programme (*outputs*), mais elle s'interroge plus largement sur les « résultats » pour la société (*outcomes* et *impacts*), sur la façon dont ils sont obtenus, sur le rôle des acteurs, pour apprécier la pertinence de la politique au vu des besoins de la société et de leur évolution. Cette appréciation dépend du point de vue des acteurs qui sont multiples (financeurs, opérateurs, agents de la mise en œuvre, bénéficiaires, citoyens) et le processus d'évaluation tient compte de cette complexité, des diverses manières possibles adaptées au contexte.

Les pays anglo-saxons, et plus particulièrement les États-Unis apparaissent souvent comme des pays où l'évaluation se pratique communément et depuis longtemps, accréditant l'idée d'un « retard » français. Relativisant cette idée, Vincent Spenlehauer nous invite à une plongée dans l'histoire des sciences sociales états-uniennes pour montrer comment les sciences politiques et les sciences sociales orientées vers l'analyse de politiques se sont développées dans une relation à l'État bien différente de celle de la France. Aux États Unis, les sciences sociales *policy oriented* ont contribué à construire une théorie de l'État dans un pays où la centralité de l'État ne va pas de soi et où la place et le rôle des citoyens dans une démocratie est une recherche permanente ; l'évaluation des politiques publiques s'inscrit dans ces débats scientifiques nourris, appuyés sur des recherches empiriques nombreuses qui ont contribué à l'inventivité des méthodes¹⁴. L'histoire française de l'État s'enracine dans une histoire longue bien

12. Définition de l'Institut français du contrôle et de l'audit interne (IFACI). Cf. le décret du 28 juin 2011 sur l'audit interne dans l'administration : « assurer la maîtrise des risques ». Sur l'audit interne, voir également le texte d'Hervé Boullanger, « L'audit interne dans le secteur public », ce numéro, p. 1029.

13. Dans l'expression « contrôle de gestion », le mot contrôle est un faux ami importé de l'anglais où *control* signifie « maintenir (un système en l'état) ».

14. Des plus anciennes (*pluralistic and utilization-focused evaluation*) aux plus récentes (*quasi-experimental design, theory-driven evaluation*) prônées par les instances européennes.

différente¹⁵, et ce que nous devons retenir de l'histoire américaine, ce sont les apports nombreux de cette longue pratique de l'évaluation faite par des chercheurs universitaires. La recherche en sciences sociales s'est davantage investie aux États-Unis dans l'évaluation productrice de connaissances qu'en France où, après la disparition du Conseil scientifique de l'évaluation, le lien entre l'évaluation des politiques publiques et le monde académique s'est distendu avec la disparition de grands appels d'offre centraux. Cela étant, au niveau local, de façon dispersée et peu visible, des chercheurs ont continué à être sollicités et à faire des recherches à visée évaluative auprès de commanditaires tels que des opérateurs de l'État ou des collectivités territoriales¹⁶.

CE QUE NOUS APPRENNENT LES TÉMOIGNAGES

En dépit de l'apparent effacement de l'évaluation dans le discours politique dans les années 2000 au profit de la LOLF, puis de la RGPP, la culture de l'évaluation s'est diffusée, tant dans certaines administrations centrales (santé, éducation, emploi...) que dans les collectivités territoriales et dans les agences et les opérateurs de l'État mais sans faire de bruit. Rappelons que les politiques contractuelles portaient en elles le germe de l'évaluation : les contrats de plan État-région (CPER devenus contrats de projets État-région), les Fonds structurels, ou encore les contrats d'objectifs et de moyens entre l'État et ses opérateurs. Aussi les conseils régionaux ont-ils été les premiers parmi les collectivités territoriales à faire des évaluations sous la double contrainte des contrats de plan et de la Commission européenne qui diffuse des standards internationaux. Ayant alors acquis une culture de l'évaluation, la plupart d'entre eux s'en sont saisis pour évaluer leurs propres politiques. Ainsi, par exemple, la région Rhône-Alpes a-t-elle engagé en 2012 deux évaluations concernant ses deux grands domaines de compétence que sont la politique de transports et la politique en faveur des lycées, soit les trois quarts de son budget.

Dans ce dossier, deux textes témoignent de ces pratiques, une collectivité territoriale (un département) et une agence (Pôle Emploi). Responsable de l'évaluation au Conseil général des Côtes d'Armor, Gwendoline Lions décrit comment l'évaluation, démarche relativement récente portée par l'exécutif, a pu s'intégrer progressivement dans les directions opérationnelles ; elle analyse les facteurs qui ont permis d'inscrire la dimension évaluative au sein de l'action publique locale.

Avec le recul de plusieurs années de pratiques de l'évaluation à Pôle-emploi, François Aventur, sous-directeur de la stratégie et des évaluations, pose la question du lien entre évaluation et décision. Alors que ce lien est inscrit dans l'organigramme (le « comité d'évaluation » a été créé au sein du conseil d'administration de Pôle-emploi), alors que ces évaluations suivent des protocoles bien formalisés (depuis l'élaboration du programme annuel d'évaluation, la mise en place d'instances dévaluation et de comités

15. En France, la souveraineté nationale est l'héritière directe de la souveraineté monarchique et décrète l'intérêt général dans une vision surplombante et généralement perçue comme protectrice, comme l'était le Roi, père de ses sujets (Rosanvallon 2006). La notion d'intérêt général se décline différemment : en France les institutions en sont le garant dans une vision descendante (*top down*) ; aux USA, le « bien commun » se construit localement d'abord par la composition des intérêts particuliers (*bottom up*) (Thévenot 2001).

16. Ainsi exemples parmi d'autres : l'université de Créteil pour le Conseil général de l'Essonne sur la garde des jeunes enfants ou sur la prévention spécialisée ; ou l'université de Paris XIII pour le Conseil général de la Seine Saint Denis sur des ateliers coopératifs ; ou encore l'École d'économie de Paris, pour Pôle emploi sur l'accompagnement renforcé des demandeurs d'emploi, et pour l'Observatoire national des zones urbaines sensibles sur les programmes de réussite éducatives, etc.

de pilotage, jusqu'à la présentation au conseil d'administration et à la publication des résultats), en dépit de ces facteurs favorables, le lien avec la décision reste indirect. L'articulation entre évaluation et décision relève avant tout d'un effet de « capillarité » ou de diffusion des connaissances produites par l'évaluation parmi les décideurs, gestionnaires et techniciens de l'offre de service.

La pratique de l'évaluation se développant, des acteurs se spécialisent : une nouvelle profession ou des nouvelles professions sont-elles en train d'apparaître et de se constituer ? Nicolas Matyjasik, chercheur à l'Institut de la gestion publique et du développement économique, décrit un processus de professionnalisation en cours, que ce soit du côté des prestataires privés ou des agents administratifs qui commanditent et suivent les évaluations. Ces deux mondes s'interpénètrent puisque les consultants privés qui font de l'évaluation ont souvent eu des parcours hybrides, avec une certaine « appétence » pour la chose publique. Au-delà d'un apprentissage sur le tas, réutilisant des savoirs existants comme le management de projet, et d'une lecture des guides de bonnes pratiques¹⁷, des masters universitaires se créent, premiers pas vers la reconnaissance d'une profession, comme en atteste déjà l'existence de fiches de poste « chargé(e) d'évaluation de politiques publiques » dans le répertoire des métiers du centre national de la fonction publique territoriale¹⁸. Mais le chemin reste long.

LA RELANCE DU DÉBAT SCIENTIFIQUE SUR L'ÉVALUATION

Le renouveau constaté concerne également les méthodes utilisées. À la fin des années 2000, la demande politique a rencontré l'actualité scientifique avec, d'un côté, l'accent mis sur « le jugement par les résultats », tant par le gouvernement français que par la Commission européenne, de l'autre la venue en France d'Esther Duflo, du Massachusetts Institute of Technology, invitée au Collège de France à la chaire « Savoirs contre pauvreté » durant l'année scolaire 2008-2009 (Duflo, 2009). Cette dernière a diffusé à une large audience le recours à la méthode de l'expérimentation aléatoire contrôlée (*RCT* : *randomized controlled trial*), pour juger de l'efficacité d'un dispositif.

Cette méthode consiste à comparer une population « traitée » (à l'instar d'un test de médicament), c'est-à-dire qui a bénéficié du dispositif, à une population identique sous tous rapports qui en a été exclue (groupe témoin, ou groupe de contrôle). Jusqu'à la réforme constitutionnelle de 2003¹⁹ qui a autorisé les expérimentations (sous condition de généralisation future), celles-ci étaient jusque-là interdites en France au nom de l'égalité de traitement des citoyens.

Cette triple conjonction, politique, scientifique et législative, a permis de lancer des chantiers nouveaux d'évaluation avec groupes témoins ; ce furent les expérimentations lancées par Martin Hirsch, d'abord sur le Revenu de solidarité active généralisé

17. Voir sur le site de la SFE (www.sfe-asso.fr) les nombreux guides existants, généralistes ou spécifiques à un domaine.

18. Catégorie des métiers de direction générale Fiche n° 05/A/01 – « chargée / chargé d'évaluation des politiques publiques »

19. La loi constitutionnelle n° 2003-276 du 28 mars 2003 relative à l'organisation décentralisée de la République (NOR: JUSX0200146L03) consacre par deux fois le principe de l'expérimentation : 1) le nouvel article 37-1 autorise la loi ou le règlement à comporter, pour un objet et une durée limitée, des dispositions à caractère expérimental ; 2) le nouvel alinéa 4 de l'article 72 en fait de même pour les collectivités territoriales dans le cadre d'une loi organique, sous condition de généralisation. C'est l'objet de la loi organique n° 2003-704 du 1 août 2003 relative à l'expérimentation par les collectivités territoriales (NOR: INTX0300039L).

dès 2008 avant la fin de l'évaluation²⁰, puis en tant que Haut-Commissaire à la jeunesse, imposées dans l'appel à projets du Fonds « expérimentation jeunesse » où, pour être retenus, les projets devaient impérativement prévoir et intégrer une démarche d'évaluation par comparaison de groupes témoins²¹. A également fait figure de chantier pionnier par son ampleur, l'évaluation de l'accompagnement renforcé des demandeurs d'emploi lancée en 2007 (Behagel *et al.* 2009), puis celle lancée par Pôle-emploi en 2009 (DARES 2013), *etc.*

L'engouement pour cette méthode, considérée par certains prosélytes comme le *golden standard* (étalon-or) de l'évaluation donnant la « preuve par l'évidence » du succès ou de l'échec de la politique (*evidence-based policy*), a suscité bien des débats, tant sur le plan éthique que scientifique (réactivant d'ailleurs des débats déjà anciens au sein de la communauté internationale des évaluateurs, notamment outre-atlantique). Sur un plan pratique, la construction de groupes témoins ne peut pas toujours être réalisée de manière aléatoire mais doit tenir compte des réalités sociopolitiques de mise en œuvre²². Exclure des bénéficiaires potentiels au nom d'une méthode « scientifique » pose des problèmes éthiques : certains agents opérationnels de Pôle-emploi par exemple ont mal supporté que ce soit le tirage au sort par la machine qui leur dicte l'orientation vers tel type d'accompagnement, disqualifiant leur compétence et leur honneur professionnel.

Sur le plan scientifique, la méthode *Randomized controlled trial* suppose que la loi des grands nombres et l'hypothèse « toutes choses étant égales par ailleurs » soient vérifiées ; elle n'est donc applicable qu'à des dispositifs relativement simples, homogènes et stabilisés, et sur des populations nombreuses ; dans ces cas, elle mesure l'existence ou non d'un résultat, mais ne décrit pas comment il est obtenu. L'analyse de situations complexes ou la compréhension des processus d'action requièrent d'autres méthodes (SFE 2011).

De fait, « la preuve par l'évidence » que les méthodes expérimentales ou quasi-expérimentales (avec groupes de contrôle) sont supposées apporter n'empêche pas les controverses. Un bel exemple de telles controverses nous est donné par Arnaud Lacheret à propos du « chèque éducation » (*school vouchers*) expérimenté aux États-Unis, dont les études à visée évaluative multiples et réitérées pour légitimer son éventuelle extension au niveau fédéral ont fait l'objet de débats, riches et parfois très vifs, au sein de la communauté scientifique américaine, entre 1998 et 2001, avec des conclusions en partie contradictoires... L'article montre combien l'application de la méthode expérimentale a du mal (ou ne peut pas) appréhender tous les facteurs qui jouent sur le choix d'une forme de scolarisation ; politologues, sociologues ou spécialistes des sciences de l'éducation privilégiant tel ou tel facteur explicatif. En conséquence, en dépit d'un questionnement évaluatif ciblé, les « évidences » des uns n'en sont pas pour d'autres car elles sont tributaires des hypothèses et paradigmes retenus. De la critique sur les méthodes, l'article passe alors à une critique de l'usage qui est fait de l'évaluation pour éviter le débat politique sur les valeurs qui sous-tendent le dispositif et le transformer en une discussion technique.

20. Voir le *Rapport final* du Comité d'évaluation du RSA (2009), présidé par François Bourguignon (travaux réalisés par la Cnaf, la Drees, l'EEP).

21. Le Fonds doté de 150 millions d'euros a retenu et financé 167 projets sur les quelques 506 qui lui ont été soumis (au 4 mai 2012). Voir sur le site du CEREQ les 27 expérimentations que cet organisme a évaluées.

22. Dans le cas du RSA, le biais de sélection illustre la tension entre un objectif scientifique et un objectif de politiques publiques : volonté d'une part des économistes de choisir les départements au hasard pour éviter tout biais dans l'évaluation et volonté d'autre part des conseils généraux de maximiser l'efficacité et/ou l'équité dans la lutte contre la pauvreté.

REGARDS CRITIQUES

Transformer un débat politique en une discussion technique n'est qu'une des critiques faites à l'évaluation. En France, la relance de l'évaluation tous azimuts et l'ambiguïté d'un discours politique qui mêlait évaluation d'actions collectives et sanction individuelle²³, ont justifié bien des analyses critiques (Garcia, Montagne 2011). Ce dossier illustre trois familles de critiques traitées ensemble dans les trois articles suivants.

La première série de critiques, énoncée dans chacun des articles, est celle de la croyance en la rationalité de l'action publique, que porte en elle une certaine vulgate de l'évaluation qui consiste à comparer mécaniquement des objectifs à des résultats. Or, une politique publique n'a pas toujours des objectifs clairement énoncés (l'action politique n'est-elle pas justement l'art du compromis, qui souvent n'existe que par l'accord sur des ambiguïtés ?). Ensuite, la vision mécaniste d'une action entraînant linéairement et causalement des effets méconnaît la complexité du réel, les problèmes d'imputabilité, les jeux d'acteurs, l'évolution des contextes et des besoins, le rôle des perceptions subjectives de l'action.

La deuxième famille de critiques s'attaque au rôle des indicateurs et, plus généralement, à l'usage de la mesure, qui gomme la complexité du réel, ne donne à voir que les dimensions quantifiables et qui permet un contrôle techniciste peu informé (Data 2009). La critique des indicateurs de la LOLF n'est plus à faire (Brunetière 2006, 2010). Cette critique est d'autant plus forte que l'on passe sans ambages de l'évaluation collective à l'évaluation individuelle, méconnaissant la rationalité intrinsèque des acteurs et leur subjectivité (Del Rey 2013).

C'est l'objet de la troisième famille de critiques illustrée ici par Lucie Mottier qui s'appuie sur l'histoire ancienne de la mesure dans le domaine de l'éducation, et par Michel Chauvière qui prend le point de vue du travail professionnel, notamment du travail social, quand les pratiques évaluatives, en supprimant l'incertitude et la subjectivité de l'humain, tendent à disqualifier les compétences créatives et analytiques des personnes, portant atteinte à leurs identités professionnelles et à l'honneur de leurs fonctions et professions.

À ces critiques, d'aucuns répondent que l'une des vertus de l'évaluation, quand elle ne se confond pas avec le contrôle-sanction, quand elle est menée comme un processus sur des actions collectives tenant compte des acteurs, est justement de « faire sortir la pépite du désaccord de la gangue du malentendu » (Viveret, 2006) ; tout l'art de l'évaluation serait alors de créer un processus dynamique vertueux, porteur « d'intelligence collective »²⁴.

Lucie Mottier, après avoir rappelé l'origine de la mesure en éducation avec la docimologie qui en définit le modèle de référence, propose un autre modèle fondé sur la nature socialement située de l'évaluation (les différents acteurs, gestionnaires, pédagogues, ont des objectifs et des valeurs différentes) ; il s'agirait de trouver des formes où se rencontreraient et se composeraient ces différentes cultures, réconciliant évaluation individuelle et action collective.

De façon parallèle, Jean-René Brunetière remarque que bien des échecs dans l'usage des indicateurs dits de performance proviennent de la différence fondamentale de nature entre ce qu'on appelle « objectif » en politique et en gestion. Dès lors, le succès des politiques publiques supposerait la construction, par l'ensemble de l'appareil d'État avec ses partenaires, de plages de consensus et de compromis sur le sens de l'action à mener, pour définir ensemble la « performance sociale ».

23. Cf. discours des vœux du Président de la République à l'Université et la Recherche (janvier 2009).

24. L'« intelligence collective », terme souvent utilisé par Michel Rocard à propos de l'évaluation, est fille de l'*organized intelligence* de Lippman (1922), citée par Vincent Spenehauer dans ce dossier.

Ces critiques soulignent la nécessité, et la difficulté, d'articuler le modèle de décision rationnelle qui sous-tend l'ambition « positiviste » de l'évaluation, et la dimension « réflexive » de celle-ci, qui fait droit à la complexité, au caractère mouvant des objectifs et à l'autonomie des acteurs.

OUVRIR LE CHAMP DES CRITÈRES D'ÉVALUATION

Les « bonnes pratiques » de l'évaluation ne sont pas figées, et des travaux de recherche explorent de nouvelles pistes pour améliorer la qualité des évaluations et de leurs processus. La troisième partie de ce dossier commence par un témoignage. Didier Houssin, devenu président de l'Agence nationale de l'évaluation de la recherche et de l'enseignement supérieur (AERES)²⁵ en 2011, après avoir soutenu l'évaluation des politiques de santé dans sa fonction précédente de directeur général de la santé²⁶, analyse l'usage de la coproduction pour améliorer la qualité et la légitimité des évaluations.

Puis trois articles tracent des perspectives qui ouvrent le champ des critères d'évaluation. Le premier – Hélène Milet et Philippe Warin – concerne la prise en compte des publics, ceux de la politique menée et ceux sollicités dans le processus d'évaluation ; les auteurs observant que des bénéficiaires potentiels ne recourent pas à des dispositifs pourtant créés pour eux ; ils proposent que ce phénomène de « non-recours » devienne un critère pour analyser la pertinence de la politique. Le deuxième article – Christine Offredi et Michel Laffut – se penche sur la question du bien-être qui est l'objectif final de toute politique ; à ce titre c'est un critère potentiel d'évaluation à condition de le rendre opérationnel ; les auteurs explorent le concept et analysent deux expérimentations qui tentent de le mettre en œuvre empiriquement. Le troisième article – Maurice Baslé – plaide pour une analyse renforcée de la qualité des évaluations ou « méta évaluation », pratique encore peu fréquente en France.

*

* *

Ce dossier thématique – qui peut apparaître comme une mosaïque « d'impressions » sur l'évaluation, ses pratiques et défis – n'épuise pas le sujet : des interrogations subsistent. Des acteurs, de plus en plus nombreux, procèdent à des évaluations des dispositifs existant dans l'idée d'optimiser le service rendu par la dépense publique. Ces pratiques sont dispersées, peu visibles (leur apport spécifique à l'amélioration de la gouvernance publique est encore mal compris par les décideurs et les médias). Elles mobilisent peu la recherche, sinon de façon marginale ou rarement revendiquée. Les connaissances produites par ces évaluations sont peu capitalisées, en l'absence d'une discipline académique reconnue. Visiblement, une profession se cherche, mais est-elle vraiment en cours de constitution ?

Ces interrogations sont aussi celles de la Société française de l'évaluation, dont le développement reflète celui des pratiques décrites dans les différents articles. Association à but non lucratif, elle rassemble les diverses personnes et organismes qui s'intéressent à l'évaluation

25. Agence nationale de l'évaluation de la recherche et de l'enseignement supérieur, transformée en Haut Conseil de l'évaluation de l'enseignement supérieur et de la recherche, par la loi n° 2013-660 du 22 juillet 2013 (art. 90), relative à l'enseignement supérieur et la recherche.

26. Évaluations portant sur le plan cancer, la périnatalité, les maladies orphelines et évaluations réalisées dans le cadre du Haut Conseil de la Santé Publique.

des politiques publiques ; à sa création en 1999, lorsque le dispositif central s'étiolait, ses membres se répartissaient en trois tiers à peu près égaux de chercheurs, de commanditaires administratifs et de prestataires privés. Aujourd'hui, son audience s'est fortement accrue parmi les collectivités territoriales, les opérateurs de l'État et les gestionnaires d'action publique (associations d'action sociale, notamment), et les prestataires privés ; le poids de ces catégories s'est accru au détriment de celle des chercheurs et universitaires, dont le rattachement principal est davantage la discipline d'origine que le thème de l'évaluation²⁷.

Les chercheurs qui travaillent sur les différentes formes d'évaluation sont dispersés dans plusieurs disciplines académiques au gré des sujets étudiés. Le caractère nécessairement interdisciplinaire des compétences mobilisées par les évaluations (management de projet, recueil et traitement de données qualitatives et quantitatives, selon des méthodes propres à chaque discipline des sciences sociales : sociologie, économie, statistique, économétrie, anthropologie, communication, sciences politiques...) rend difficile l'implication de chercheurs visant une carrière académique dans un mode univoque. La dispersion des financements, la crainte (réelle ou supposée) d'être instrumentalisés, ont pu éloigner des chercheurs. Pourtant quelques indices laissent présager d'un mouvement de retour : la création du Laboratoire interdisciplinaire d'évaluation des politiques publiques (LIEPP) créé à Sciences-Po en 2012²⁸, les travaux récents de l'École d'économie de Paris pour de nombreux commanditaires et qui associent d'autres regards à l'expérimentation aléatoire contrôlée, le succès d'audience de l'université d'été internationale organisé par le réseau des chercheurs de la Société française de l'évaluation en 2012 à l'Institut d'études politiques de Lille...

Pour remédier à la dispersion des évaluations réalisées et mieux connaître la pratique de l'évaluation en France aujourd'hui, la Société française de l'évaluation, en lien avec le Secrétariat général à la modernisation de l'action publique, a monté un observatoire des évaluations réalisées. Gageons qu'à terme, cet observatoire pourra donner une vision annuelle de l'actualité des pratiques et des thèmes traités, et permettre une certaine capitalisation de ces expériences, concourant ainsi à asseoir l'évaluation des politiques publiques dans le paysage administratif et cognitif français.

Références bibliographiques

- Barbaroux, Jérôme (2007), « La dimension pluraliste de la démarche d'évaluation des politiques : de la théorie à la pratique dans la Région Rhône-Alpes » in Jacot H. et Fouquet A. (dir.) *Le citoyen, l'élu, l'expert. Pour une démarche pluraliste de l'évaluation des politiques publiques*, L'Harmattan/ SFE, 241 p.
- Behaghel, Luc ; Crepon, Bruno et Gurgand, Marc (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 économique de Paris.

27. De manière significative, aux journées de Grenoble, en juin 2013, les participants se répartissaient ainsi : 41 % venaient de collectivités territoriales, 26 % de l'État central et déconcentré, d'organismes publics et d'associations d'intérêt général, 17 % de cabinets de consultants, 14 % d'universitaires et de chercheurs et 2 % de retraités ou sans emploi.

28. Le LIEPP a reçu le label de laboratoire d'excellence (Labex) assorti de financements très importants sur la base d'une promesse dont on attend encore les débuts de réalisation, à savoir la relance de travaux de recherche interdisciplinaire sur et en évaluation.

- Brunetière, Jean-René (2006), « Les indicateurs de la LOLF, une occasion de débat démocratique ? », *Revue française d'administration publique*, n° 117.
- Brunetière, Jean-René (2010), « Les objectifs et les indicateurs de la LOLF, quatre ans après.. », *Revue française d'administration publique*, n° 135, p. 477-495.
- Comité d'évaluation des expérimentations (2009), *Rapport final sur l'évaluation des expérimentations RSA*, La documentation française.
- DARES (2013), *L'accompagnement renforcé des demandeurs d'emploi : évaluation du recours aux opérateurs privés par Pôle emploi de 2009 à 2011*, DARES Synthèses.Eval, n° 3, janvier 2013.
- DATA Lorraine (2009), *Le Grand trucage : comment le gouvernement manipule les statistiques*, Paris, coll. cahiers libres, La découverte.
- Del Rey, Angelica (2013), *La Tyrannie de l'évaluation*, Paris, La découverte.
- Duflo, Esther (2009), *Expérience, science et lutte contre la pauvreté*, Paris, Fayard.
- Fouquet, Annie (2010), « L'usage des statistiques : de l'aide à la décision à l'évaluation des politiques publiques », *Revue française des affaires sociales*, janvier-juin 2010, p. 309-322.
- Gibert, Patrick (2010), « Contrôle et évaluation, au-delà de querelle sémantiques parenté et facteurs de différences », *Revue française des affaires sociales*, janvier-juin 2010, p. 73-88.
- Garcia, Sandrine ; Montagne, Sabine (2011), « Pour une sociologie critique des dispositifs d'évaluation », *Actes de la recherche en sciences sociales*, 2011/4, n° 189.
- IGA, IGF, IGAS (octobre 2012), *Bilan de la RGPP et conditions de réussite d'une nouvelle politique de réforme de l'État*, Paris, La documentation française, 367 p.
- IGA, IGF, IGAS (décembre 2012), *Cadrage méthodologique de l'évaluation des politiques publiques partenariales*, Guide, Paris, La documentation française, 54 p.
- Mouterde, François et Trosa, Sylvie (2010), *Les nouvelles frontières de l'évaluation, 1989-2009 : Vingt ans d'évaluation des politiques publiques en France, et demain ?*, L'Harmattan/SFE, 339 p.
- Rosanvallon, Pierre (2006), *Le modèle politique français : La société civile contre le jacobinisme de 1789 à nos jours*, Paris, Le seuil.
- Société française de l'évaluation (2006), *La Charte de l'évaluation en France*, http://www.sfe-asso.fr/sfe-evaluation.php?menu_id=191.
- Société française de l'évaluation (2009), *Voisinages : de la confusion à la complémentarité. Évaluation, audit de performance, pilotage de la performance, management de la qualité*, Cahiers de la SFE n° 4 (en partenariat avec France qualité publique et l'Afigese).
- Société française de l'évaluation (2011), *Évaluation des impacts des programmes et services publics*, Cahiers de la SFE n° 6.
- Thevenot, Laurent (2001), « S'associer pour composer une chose publique », in Chapart J.-N. et alii, *Actions associatives, solidarités et territoires*, Saint-Étienne, Publications de l'Université de Saint-Étienne, p. 267-274.
- Viveret, Patrick (2006), « Qualité démocratique et construction des désaccords », in Caille A. (dir.), *Quelle démocratie voulons-nous ?*, Paris, La découverte, p. 32-34.

