

La fièvre de l'évaluation

L'évaluation est devenue omniprésente, qu'il s'agisse de recherche ou de soins, d'individus ou de lois. Célébrée ou critiquée, elle est érigée en priorité des politiques publiques.

JULIEN DAMON

Une culture de l'évaluation ?

Qu'ont en commun l'enfant en maternelle, le ministre, le chercheur, le médecin ou encore le policier ? Ils peuvent aujourd'hui être « évalués ». Les évaluations, en complément et en développement des traditionnelles notations et appréciations, se sont implantées particulièrement dans l'éducation. Les élèves sont évalués, mais aussi les enseignants nationaux. Nombre de controverses techniques et de discussions de parents étonnés accompagnent cette généralisation de l'évaluation des élèves. Des polémiques tout aussi nourries accompagnent les discussions sur les évaluations des enseignants-chercheurs (*article*, p. 22) comme celles sur les programmes éducatifs et leurs résultats (*encadré* p. 23).

Depuis une trentaine d'années, l'évaluation s'est affirmée dans le cadre d'un puissant mouvement de réforme des politiques publiques. Après les risques et les investissements, on évalue maintenant les hôpitaux, les chercheurs, les universités, les services de sécurité, les psychothérapies. De l'évaluation individuelle des aptitudes et des compétences à l'évaluation par les parlementaires, les experts et les usagers de politiques sanitaires, tout est devenu objet potentiel d'évaluation.

On est même passé à une « culture de l'évaluation » répandue dans de nombreux domaines. Si le principe de l'évaluation (consistant à mettre en évidence la valeur d'un service ou d'une relation) n'est généralement pas contesté, la dérive gestion-

naire qui peut l'incarner est très souvent rejetée. Il s'ensuit qu'évaluation rime autant avec rationalisation (attendue par les évaluateurs) qu'avec exaspération (vécue par les évalués). Cette affirmation progressive de l'évaluation suscite les espoirs de ceux qui en attendent une amélioration des activités humaines. Elle soulève également des critiques, voire des disputes virulentes, car tout ne serait pas évaluable. Au fond, la question de l'évaluation revient à une problématique assez classique. Elle s'incarne en une formule connue, sans paternité bien établie : ce qui a de la valeur n'a pas toujours de prix, et ce qui a un prix n'a pas nécessairement de valeur. Une des premières tensions de l'évaluation est en effet son penchant pour la quantification. ■



Rob Colvin/Getty Images

Qu'est-ce qu'évaluer ?

Pour les économistes ou les comptables, évaluer c'est d'abord estimer, sans nécessaire appréciation politique. Pour les ingénieurs, l'évaluation c'est la mesure, d'essence technique, physique et mathématique. Pour les professeurs, évaluer c'est émettre une appréciation, souvent incarnée par la note. Or l'évaluation ne saurait se limiter à la notation.

L'opération d'évaluation consiste, financièrement, à calculer une valeur d'après des données et des critères déterminés. Concrètement, évaluer c'est d'abord chiffrer ce que vaut en argent un bien, un avantage ou un dommage. Terme maintenant plus général qu'estimation ou appréciation, l'évaluation n'est plus seulement une seule mesure de coût, de résultat ou de conformité. C'est un ensemble de procédés pour que des évaluateurs puissent formuler une opinion, voire un jugement, sur l'adéquation d'un programme, d'une politique, d'une activité, avec les objectifs qui leur sont fixés.

L'évaluation n'est pas une discipline avec ses approches et ses méthodes formatées : elle a pour visée essentielle d'être un outil d'aide à la décision. Sous couvert d'objectivité voire de scientificité, elle est d'abord un instrument de gestion. ■

Une idéologie libérale ?

L'institutionnalisation progressive de l'évaluation, surtout conçue comme un élément de la performance publique, est souvent analysée comme l'introduction en France des théories et des outils du *new public management*. Ce mouvement composite rassemble un ensemble de réformes menées depuis trente ou quarante ans et habilement promues par les Anglo-saxons. Il s'agit en premier lieu d'améliorer la qualité des services publics, avec la nécessité de tenir compte de la sensibilité croissante des citoyens au thème du *value for money* (rapport qualité/prix). Mais il y a également la volonté de rendre plus transparente l'action administrative avec un souci d'*accountability* (les Québécois traduisent parfois par « redevabilité ») consistant à renforcer le contrôle démocratique de la gestion publique. Cette tendance suscite des réactions très contrastées. Pour certains, l'évaluation représente l'alpha et l'oméga de la modernisation de la dépense publique. Mais pour d'autres, elle représente le dernier avatar de la pensée (dite ultra) libérale, car elle revient – ce qui est vrai – à transposer dans le secteur public des instruments de gestion qui prévalent dans le privé. Ce qu'ils dénoncent comme une illusion, car, en un sens, la production publique serait inévaluable. Entre les deux, certains considèrent qu'elle a sa pertinence mais qu'elle ne peut s'exercer dans tous les domaines, en particulier ceux où l'essentiel relève de la qualité de la relation humaine (travail social ou soins). ■ J.D.

Quand est-elle apparue ?

Les racines intellectuelles de l'évaluation proviennent de la gestion. Comme ancêtre de l'évaluation, on cite souvent les systèmes américains de *planning programming budgeting* (PPB), établis dans les années 1960. Ils ont servi aux expériences de rationalisation des choix budgétaires (RCB) en France dans les années 1970. Cela consiste à réaliser des analyses sur les coûts et sur l'efficacité des programmes publics. Mais ce n'est qu'à partir des années 1980 que l'on se met à parler véritablement d'évaluation. En 1989 sont ainsi mis en place un Comité national d'évaluation de la recherche ainsi qu'une Commission nationale d'évaluation du RMI. Un Conseil supérieur de l'évaluation (CSE) est créé en 1990. Il sera remplacé par

un Conseil national de l'évaluation (CNE) en 1998, aujourd'hui disparu.

Institutionnellement, un changement important est introduit avec la loi organique relative aux lois de finances (LOLF) votée en 2001, qui engage profondément la France dans un mouvement de réforme affectant le cœur de l'administration. Cette loi bouleverse les règles de présentation, d'adoption et d'exécution du budget de l'État. Sa mise en œuvre progressive doit grandement affecter les modes de fonctionnement et d'évaluation des services rendus à tous les publics. D'une obligation de moyens, l'État passe désormais avant tout à une obligation de résultats. Le budget de l'État est maintenant décomposé en missions, en programmes et en

actions, chacun de ces registres étant doté d'objectifs, très généralement quantifiés. Dans cette logique de mesure de l'efficacité publique, la révision générale des politiques publiques (RGPP) est lancée en 2007 (1). Conduite par la haute fonction publique, elle consiste en une évaluation à vocation très large de l'ensemble des interventions des pouvoirs publics. Cette RGPP toujours en cours est gérée par l'appareil administratif. La révision constitutionnelle de 2008 a enfin introduit quelques mots décisifs à l'article 24 de la Constitution. Il y est précisé que le Parlement vote la loi, qu'il contrôle l'action du gouvernement et, désormais, qu'il évalue les politiques publiques. ■

(1) www.rgpp.modernisation.gouv.fr

Qui doit évaluer ?

L'activité d'évaluation doit-elle être laissée aux experts et cabinets privés ? Ou bien relève-t-elle du rôle des élus et des représentants ? Le vote n'est-il pas, en dernier lieu, la meilleure des évaluations d'une politique ? Sous toutes ces questions se trouve le lien problématique entre évaluation et démocratie.

Pierre Rosanvallon, dans son nouveau cycle de travaux sur la démocratie, rappelle que le citoyen est à la fois électeur et juge, et que la légitimité démocratique procède à la fois de l'élection et de l'administration. Il propose un tableau de synthèse (1) où il montre que de fait, toutes les procédures d'évaluation relèvent, dans une certaine mesure, du jugement politique. Et dans une autre mesure le jugement politique, sous ses diverses formes, est une évaluation.

P. Rosanvallon, dans son «*exploration de l'univers contre-démocratique*», souligne le passage d'une démocratie de confrontation et de représentation à une démocratie d'imputation. Les gouvernants ne doivent pas seulement présenter puis mettre en œuvre des programmes et des



Mettre son bulletin de vote dans l'urne est aussi une forme d'évaluation.

idées, ils doivent rendre des comptes et être à l'écoute. Ils sont évalués sur leur légitimité, leurs capacités et leurs performances ; et celles-ci sont davantage mises en question et en cause. La défiance, qui

est une des sources de la montée en puissance du besoin d'évaluation, autorise des pressions et permet des corrections. ■

(1) Pierre Rosanvallon, *La Contre-démocratie. La politique à l'âge de la défiance*, Seuil, 2006.

L'évaluation, pour quoi faire ?

L'ancien Conseil supérieur de l'évaluation distinguait trois fonctions de l'évaluation.

■ Sur un plan « cognitif », l'évaluation consiste à construire, à acquérir et à partager des connaissances. Elle doit autoriser une meilleure intelligence des problèmes traités.

■ Sur un plan « normatif », l'évaluation a pour ambition de fonder des jugements et de préconiser des choix. Elle doit être un outil d'aide à la décision.

■ Sur un plan « instrumental », l'évaluation doit être à l'origine de transformations et d'améliorations. Elle ne se comprend qu'avec des préconisations pour des décisions plus rigoureuses et un meilleur agencement des ressources.

Visant une connaissance améliorée, une communication plus adaptée, une conduite des politiques plus ajustée, elle peut en réalité relever de trois usages.

• Usage réglementaire : on fait de l'évaluation car c'est marqué dans le règlement...

Il y a là une évaluation « rituelle » sans nécessairement grande implication, ni grande utilité.

■ Usage managérial : les conclusions de l'évaluation peuvent devenir des aides à la décision.

■ Usage social : les résultats de l'évaluation sont publics et visent à être discutés dans le débat social.

Une problématique importante et sensible est celle de l'évaluation de l'évaluation. Comment en effet « évaluer » les fondements, les méthodes et les résultats de cette activité ou de cette posture qui prétend justement apporter de l'information, de la confrontation et de l'inflexion pour toute activité ? Ce qui semble assuré c'est qu'on ne peut valablement évaluer que lorsque des objectifs, considérés comme légitimes ou non, ont été fixés. En un mot, sans fixation d'objectifs, on ne voit pas comment mesurer des résultats. La grande question dès lors est celle de la pertinence des objectifs et des indicateurs. ■

Évaluons-nous les uns les autres

De multiples sites Internet permettent de noter, comparer, évaluer. Le potentiel « participatif » de la toile peut être mis au service du pire comme du meilleur. Il en va ainsi pour les clientèles du privé et du public, où des innovations récentes vont dans le sens d'une évaluation directe par les usagers, pour les usagers.

On assiste ainsi en France à l'émergence de sites d'évaluation des *people*, des capitales et des maisons de retraite. Des initiatives comme www.note2be.fr et www.note2bib.fr (évaluation des professeurs et des médecins) ont défrayé la chronique et ont fermé.

Aux États-Unis, leurs équivalents www.ratemyprofessors.com et www.ratemydoctor.com fonctionnent encore.

Confidentialité, respect de la vie privée, qualité des données ne sont pas forcément assurés. Mais le mouvement est lancé.

Le secteur public est également de la partie. www.hospitalcompare.hhs.gov permet la comparaison des services hospitaliers. En Californie, www.schoolfinder.ca.gov veut guider les parents dans le choix des écoles publiques. ■ J.D.

Peut-elle engendrer des effets pervers ?

L'institutionnalisation de l'évaluation conduit à la généralisation des indicateurs de « résultats ». Cette extension est souvent rejetée en tant qu'établissement d'une dictature de l'indicateur, induisant une bureaucratie de contrôle inutile. Elle donne lieu à une augmentation des tâches, à une abstraction et à une bureaucratisation des procédures. Les enseignants doivent de la sorte remplir des documents toujours plus nombreux et détaillés sur les acquis des élèves, au détriment du temps consacré à être auprès d'eux. L'évaluation sur indicateurs officiels conduit à une politique du chiffre dans la police notamment, où il s'agit d'atteindre des résultats sur certaines infractions, au détriment d'objectifs

plus généraux mais moins quantifiables comme le maintien du lien social.

L'évaluation change aussi le sens du travail de l'évalué (qui va chercher à correspondre aux attentes de l'évaluateur), voire de l'évaluateur (qui se voit confier de nouvelles responsabilités d'évaluation à côté de son premier métier). Forcés de rendre des comptes détaillés, tous les employés, du public comme du privé, voient leurs activités quotidiennes davantage façonnées et dirigées en fonction des canons de l'évaluation, et non plus nécessairement en fonction des objectifs premiers de l'institution dans laquelle ils exercent. À trop chercher la mesure et la croissance de la productivité, l'évaluation peut conduire à une baisse de production

et d'implication.

Le risque de dictature de l'indicateur est donc très élevé, rendant l'évaluation impossible si personne n'adhère à sa démarche et à son ambition. Il y a cependant une possibilité d'usages démocratiques de l'indicateur lorsque celui-ci est discuté, partagé et approprié. Il n'en reste pas moins que l'évaluation relève davantage du pouvoir que du savoir. Il n'y a pas d'évaluation pure, détachée des contingences et parfaitement objective. De l'autre côté, il ne peut plus y avoir de progrès mesuré sans évaluation. Bien comprise et bien délimitée dans ses ambitions, l'évaluation est célébrée. Trop obscure et rattachée à un unique souci de rationnement, l'évaluation est décriée. ■

Question de méthode

L'évaluation n'est pas une démarche uniforme. Par construction, elle doit s'adapter à son objet. Pour autant il est possible de mettre en évidence deux sujets, et trois moments possibles d'évaluation.

Au titre des sujets, peuvent ainsi être clairement distinguées :

● **Les évaluations de processus :** les procédures ont-elles été suivies ? Les moyens ont-ils été pertinemment affectés ?

● **Les évaluations d'impact :** le programme a-t-il permis les résultats prévus ? L'action a-t-elle eu les conséquences désirées ?

Au titre des moments, trois étapes peuvent être le thème d'une évaluation :

● **Ex ante :** les évaluations *ex ante* sont des études préalables d'impact.

L'évaluation passe par la prévision, la projection, la prospective, l'expérimentation. Un cas typique est celui de la mise en œuvre du revenu de solidarité active (RSA). Certes l'expérimentation n'a pas été menée à son terme et l'évaluation était plutôt mitigée. Il n'en reste pas moins qu'il s'agit là d'une expérience inédite et de large ampleur.

● **In itinere :** les évaluations menées au cours d'une réalisation consistent

en suivi, tableaux de bord, contrôle de gestion. Pour les entreprises c'est du simple suivi d'activité. Pour les enfants c'est le suivi de la scolarité, qui va maintenant d'appréciations très tôt dans le cursus scolaire à des notations d'examen ou de concours.

● **Ex post :** les évaluations finales sont réalisées au terme d'un programme ou d'une action. De fait ces évaluations sont les plus rares et les plus discutées. Il faut en effet décider que ce qui est à évaluer est terminé. Or à quel moment s'arrête une relation de travail social ou de soin ? Quand s'achève vraiment une politique publique ? ■

J.D.

Quand évaluation engendre protestation

Ces dernières années, plusieurs rapports et enquêtes à visée évaluative ont été l'objet de vives controverses.

■ PSYCHOTHÉRAPIES : le rapport censuré

En février 2004, l'Institut national de la santé et de la recherche médicale (Inserm) publie une expertise collective sur les psychothérapies. Trois approches sont comparées : les thérapies « psychodynamiques » ayant la théorie psychanalytique comme référence, les thérapies cognitivocomportementales et les thérapies « familiale et de couple ». Le rapport met globalement en évidence, sur toute une série de troubles mentaux (schizophrénie, troubles anxieux, autisme, hyperactivité...), une efficacité supérieure des thérapies cognitivocomportementales et, dans une moindre mesure, des thérapies familiales. Selon ses auteurs, l'efficacité des thérapies psychodynamiques, peu étudiée, est rarement prouvée. Les réactions à cette évaluation sont très vives, en particulier dans les milieux psychanalytiques. Si certains refusent le principe même de l'évaluation en santé mentale, d'autres pointent des méthodes biaisées. Le champ de l'expertise est ainsi limité aux maladies (alors que la santé mentale dépasse largement ce cadre). Elle laisse par ailleurs totalement de côté les consultations en cabinet

Peut-on évaluer la recherche ?

La question de la mesure des « performances » des chercheurs et des laboratoires était en bonne place des points de désaccord qui ont mené au blocage récent des universités. Des outils existent. La bibliométrie, par exemple, consiste à recenser les occurrences de citation des travaux des chercheurs et à les rapporter à la qualité scientifique des revues dans lesquelles ils sont publiés et/ou cités. Pour les détracteurs de l'évaluation, l'intrusion d'une logique de mesure chiffrée des résultats n'a aucun sens. Elle nuit à la liberté et à la créativité nécessaires à toute découverte. Nombreux sont ceux qui rappellent également qu'en dehors de la bibliométrie, l'activité des chercheurs

est déjà largement évaluée, dès qu'ils soumettent un article à une revue (comité de lecture), qu'ils candidatent à un poste (jury de sélection), sollicitent des crédits de recherche...

Il est certain que l'évaluation en s'institutionnalisant se bureaucratise nécessairement, par exemple avec des institutions dédiées (et dès lors contestées par ceux qui n'en reconnaissent pas la légitimité), comme l'Agence d'évaluation de la recherche et de l'enseignement supérieur (Aeres). Entre le bureaucrate pur qui attend tout du sceau de l'évaluation et le scientifique pur qui craint le joug de l'évaluateur, la vertu est de chercher à objectiver des activités, de manière posée et équilibrée. ■



James Yang/Getty Images

privé. Enfin, le fait de prendre comme indicateur l'amélioration de l'état du malade (par rapport à un autre malade n'ayant pas suivi de thérapie) empêche de voir que, bien souvent, la prise en charge thérapeutique du patient n'apporte pas d'amélioration mais empêche que son état n'empire. En février 2005, invité au Forum des psys, le ministre de la Santé Philippe Douste-Blazy affirme devant une audience de professionnels de la santé mentale que « *le premier devoir d'une société est de reconnaître qu'il n'existe pas une seule réponse à la souffrance psychique* », souffrance qui n'est « *ni évaluable ni mesurable* ». Il annonce également à son auditoire qu'il a fait retirer le rapport de l'Inserm du site internet du ministère de la Santé et qu'ils « *n'en entendront plus parler* ». L'Inserm, les auteurs de l'expertise, ainsi que de nombreuses associations de professionnels et de patients en santé mentale protesteront en vain contre cette censure. ■

■ ÉDUCATION: des comparaisons internationales réductrices ou révélatrices ?

Les enquêtes Pisa menées par l'Organisation de coopération et de développement économique (OCDE), comparent régulièrement les performances des élèves de 15 ans dans une trentaine de pays. Elles ont été fortement critiquées : logique du classement extrêmement réductrice, compétences scolaires mesurées en dehors de toute référence au contexte culturel... Des sociologues de l'éducation réputés (François Dubet, Christian Baudelot et Roger Establet...) ont néanmoins défendu la qualité de ces études, qui mettent particulièrement en évidence un trait français : un écart très grand entre les (très bonnes) performances de ses élites et celles (médiocres) d'une large frange de la jeunesse qui se retrouve rapidement en échec. ■

■ RSA: une expérimentation avortée

C'était une innovation : avant d'être adopté, le revenu de solidarité active a été expérimenté dans une trentaine de départements. Le but était de savoir s'il fonctionnait, quels devaient être son montant et les dispositifs d'accompagnement à prévoir... Des critiques méthodologiques ont été formulées (sur le choix des départements en particulier). Mais surtout l'expérience a été interrompue au bout d'à peine un an... par l'adoption et la généralisation anticipées du RSA en 2008. Elle aura donc été loin de livrer tous les enseignements qu'elle promettait, qui auraient permis, selon ses défenseurs, de mettre en place un dispositif optimal. ■

X.M.

B I B L I O G R A P H I E

Charte de l'évaluation des politiques publiques et des programmes publics
Société française de l'évaluation, 2006.
www.sfe.asso.fr

Gestion publique. L'État et la performance
Henri Guillaume, Guillaume Dureau, Franck Silvent, Presses de Sciences Po – Dalloz, 2002.

Voulez-vous être évalué ?
Jacques-Alain Miller, Jean-Claude Milner, Grasset, 2004.

L'Évaluation des politiques publiques
Bernard Perret, La Découverte, coll. « Repères », 2001, rééd. 2008.

« L'évaluation de la recherche en question »
Bernard Thomann, 28 mai 2009, disponible sur www.laviedesidees.fr

Évaluer les politiques publiques
Sylvie Trosa, « Les notes de benchmarking international », Institut de l'entreprise, 2003.