



**HAL**  
open science

## L'évaluation des politiques publiques

Yannick Rumpala

► **To cite this version:**

Yannick Rumpala. L'évaluation des politiques publiques. Science politique, Larcier; Bruxelles, 2023, 9782390132387. hal-04046390

**HAL Id: hal-04046390**

**<https://hal.science/hal-04046390v1>**

Submitted on 25 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Chapitre 11. L'évaluation des politiques publiques

Yannick Rumpala  
rumpala [at] unice.fr

in Christophe ROUX, Éric SAVARESE (dir.),  
*Science politique*, Bruxelles, Larcier, collection « Paradigme », 2017.

*Mars 2017*

Est-il utile et pertinent de poursuivre une politique d'exonération de cotisations sociales patronales ? Quels effets des allègements de cotisations sociales ont-ils sur l'emploi ? Les aides publiques aux entreprises sont-elles efficaces ? Faut-il revoir les aides personnelles au logement ? Comment s'assurer de l'adaptation des instruments devant favoriser les économies d'énergie ? Autant de questions, parmi de nombreuses autres, qui peuvent être à l'origine de démarches d'évaluation. Ces dernières se sont vues conférer un rôle accru dans un contexte d'exigences également accrues sur les interventions publiques, au point même de devenir une pièce importante dans la rhétorique de la « réforme » de l'État. De ce point de vue, l'évaluation n'a pas seulement à voir avec les politiques publiques (ce qui touche au « *policy making* »), mais aussi avec la politique (au sens de *politics*, dans la distinction que permet la langue anglaise) : elle est également un enjeu politique, traçable à travers son intégration au répertoire instrumental des modèles de « gouvernance ».

Comme source informative, la démarche évaluative vise à apprécier, voire vérifier, mesurer et expliquer les effets d'une action publique. L'évaluation peut porter sur une mesure, un programme (un ensemble structuré de mesures dans le cadre d'objectifs fixés et sur la base de moyens déterminés, généralement dans un calendrier défini), une politique publique (un assemblage de dispositifs et de procédures reliés par une orientation convergente). Pour ce travail, un réservoir de méthodes et d'approches est aujourd'hui disponible et continue à être développé, au croisement de différentes sources et formes d'expertises.

L'ensemble est devenu un domaine à part entière, croisant des préoccupations relevant de la science politique, mais la débordant également. Il s'agira dans ce chapitre de mieux comprendre les raisons des démarches d'évaluation, leur fonctionnement et leurs usages. Nous préciserons d'abord les enjeux de l'évaluation et verrons ainsi que l'évaluation peut aussi être prise comme un révélateur des évolutions des États, de leur positionnement, et de leurs modalités de fonctionnement et d'action (1). Nous en spécifierons ensuite les caractéristiques (2). Nous terminerons en examinant leurs modalités d'utilisation, ce qui permettra de pointer les limites et ambiguïtés de la démarche telle qu'elle a été majoritairement développée (3).

## 1) Extension d'un domaine d'expertise

### 1.1) Une approche en quête de distinction

Dans les dispositifs évaluatifs, l'attention est notamment orientée vers les résultats et effets d'une politique, d'un programme ou d'une mesure, qu'il va s'agir de mieux connaître et éventuellement comprendre. Ces effets sont-ils ceux qui étaient souhaités ? S'il y a un écart, quelles en sont les raisons ? Si cet écart apparaît problématique, des corrections ou des voies d'amélioration peuvent-elles (voire doivent-elles) alors être envisagées ?

Dans la classique grille séquentielle de Charles O. Jones (1970), sorte de schéma idéal-typique du processus constitutif d'une politique publique, l'évaluation correspond à la dernière des cinq séquences repérées, celle où sont questionnés et mis en perspective les effets, conséquences et éventuels aboutissements de cette politique. L'évaluation est censée contribuer à l'amélioration des politiques et dispositifs élaborés et promus par des autorités publiques. Outre des éléments de jugement, elle peut donner des appuis pour des propositions réajustées ou nouvelles. Elle est alors une aide potentiellement utile pour des décideurs ou des responsables, mais peut aussi avoir une valeur plus largement démocratique pour l'ensemble des citoyens.

Par contraste avec l'« analyse des politiques publiques » (ou la « sociologie de l'action publique ») qui est maintenant considérée comme une sous-discipline en science politique (Cf. chapitres précédents), l'évaluation des politiques publiques, notamment du point de vue des motivations, se situe dans une perspective différente. Certes, la première a contribué à asseoir scientifiquement la seconde (Duran, 2010a). Mais, d'un côté, la perspective est celle d'une « *analysis of policy* », tandis que de l'autre, c'est celle d'une « *analysis for policy* » (Gordon, Lewis and Young, 1977). Pour cette dernière, l'orientation est donc *a priori* plus pragmatique et opérationnelle. Dans ces deux dérivés des « *policy sciences* », le type de connaissances recherchées n'est pas le même. Dans l'« analyse des politiques publiques », les questionnements restent davantage académiques et tendanciellement moins normatifs. L'évaluation relève en revanche du savoir pratique (Duran, 2010), avec des finalités qui tendent vers l'intervention politique.

### 1.2) Enjeux des démarches évaluatives

La promotion des dispositifs et pratiques d'évaluation est une tendance qui dépasse largement le domaine de l'action publique (et qui va dorénavant, de fait, de l'individuel au collectif [Bureau, 2010]). L'évaluation apparaît toutefois comme un marqueur des transformations de ce domaine (Duran, 2010b). Sur longue durée, l'évaluation des politiques publiques est à ranger parmi les multiples expressions institutionnelles d'un vaste processus de rationalisation, si l'on adapte et prolonge la perspective d'analyse amorcée par Max Weber. Avec l'évaluation, un champ d'activités sociales, celles d'institutions publiques, est soumis à une approche de plus en plus méthodique. Examinée plus précisément, cette tendance rationalisatrice révèle en fait des traits particuliers. Si l'on se replace dans le sillage des réflexions de Jürgen Habermas (1978) sur les sociétés du « capitalisme avancé », les efforts en matière évaluative et leurs évolutions, *a fortiori* dans leur recherche d'efficacité et de performance, peuvent aussi être interprétés comme des productions réactives d'un système administratif mis en crise (baisse des rentrées fiscales, perte de moyens, etc.), parvenant plus difficilement à répondre aux impératifs en provenance du système économique et tentant de pallier les déficits de rationalité consécutifs par un surcroît de connaissances utilitaires. Le développement de l'évaluation s'inscrit dans un contexte, qui est également celui de la montée du *New Public Management* (cette « nouvelle gestion publique » largement inspirée des prescriptions appliquées au secteur privé) et des conceptions elles-mêmes axées sur l'efficacité, la

performance, etc. Son développement tient aussi à sa récupération comme composante des processus de « modernisation » de l'État et de l'action publique.

Mettre en place des politiques fondées sur des « données probantes » (« *evidence-based policies* ») passe de plus en plus pour une démarche logique (Laurent *et al.*, 2009). L'idée est de savoir ce qui marche (« *what works* »). Par les informations qu'elle permet de produire et de réinjecter dans des circuits décisionnels, l'évaluation peut être alors envisagée comme un « outil de renforcement des capacités d'action collective » (Perret, 2016) et comme une voie d'apprentissage, également collectif (Perret, 2008). Si l'on peut parler d'une « société de l'information », la sphère publique s'y insère ainsi par la production, la mise en circulation et la consommation d'une gamme particulière d'informations, dont le rassemblement à des fins utilitaires doit pouvoir rendre cumulable des connaissances sur des processus internes et externes. À l'échelle des systèmes sociaux, la circulation et l'utilisation de ces connaissances, même si c'est de manière variable comme on le verra, les rendent susceptibles de conforter ce type de réflexivité qu'Anthony Giddens (1994) a qualifié d'« institutionnelle » : en l'occurrence, des institutions s'interrogent sur elles-mêmes et reviennent réflexivement sur leurs actions, dans un processus qui est alors de nature à changer la teneur de ces actions (et qu'Anthony Giddens rattache de manière plus théorique à l'approfondissement de la « modernité »).

L'évaluation des politiques publiques et la place qui lui est accordée traduisent des enjeux rattachables aux caractéristiques d'une société complexe, où en tout cas il s'agit pour les instances publiques de pouvoir prendre en charge des problèmes qui paraissent eux-mêmes complexes et où s'affirment les besoins de disposer de capacités de pilotage. À l'évaluation correspond « un double registre de *connaissance* et de *justification* » (Duran, 2010b, p. 183). Cette connaissance s'applique à repérer et comprendre des effets qui peuvent être ceux escomptés ou d'autres moins prévus, qui peuvent être jugés positifs ou indésirables, qui peuvent paraître plus ou moins directement imputables à l'action entreprise et évaluée. Parmi les pluralités d'effets possibles d'une politique, donner plus de sens à certains, c'est conjointement permettre de les mettre en valeur, et par conséquent les stratégies qui en seraient à l'origine.

En révélant sa valeur instrumentale, l'évaluation est devenue un outil de gouvernement, prolongeant une longue histoire de conseil aux dirigeants politiques et de production d'une expertise dans cette perspective. Le recours à l'évaluation est aussi un révélateur d'une évolution des légitimités de l'action de l'État (Flückiger, 2007). Par elle s'expriment les souhaits ou les velléités non seulement de parvenir à une gestion rigoureuse mais aussi de conférer une certaine « transparence » à l'action publique. Le développement des pratiques évaluatives n'est pas sans liens avec les évolutions du cadre démocratique et, par certains aspects, participe de ce que Pierre Rosanvallon (2006) a appelé la « démocratie de surveillance », pouvant jouer comme un complément à la démocratie représentative.

### 1.3) Dynamiques, facteurs et logiques d'institutionnalisation

L'évaluation a bénéficié de ses premiers développements aux États-Unis, plus significativement à partir des années 1960, par rapprochements entre une offre provenant des sciences sociales et une demande politico-administrative (Spencehauer, 2016). Le contexte était alors marqué par une confiance dans les avancées scientifiques et celles qui étaient escomptées des sciences sociales semblaient aussi pouvoir s'appliquer à la conception des programmes publics, dans une logique espérée d'amélioration de leur fonctionnement et de leurs résultats.

L'intérêt en France a été plus tardif et plus lent à se développer. Contrôle juridique (essentiellement attaché à la régularité des actes) et contrôle de gestion (plutôt axé sur le « bon emploi » des fonds

publics) y ont en effet imprégné la culture politico-administrative, notamment à travers les interventions des corps d'inspection générale, du Conseil d'État et de la Cour des comptes. Des prémisses d'activités de type évaluatif pouvaient certes être perçues dans l'expérience de la RCB (Rationalisation des Choix Budgétaires) tentée dans les années 1960-1970 et devant permettre de réfléchir en amont sur les objectifs étatiques (Perret, 2006). Mais c'est surtout dans les années 1980, et même les années 1990, que la dynamique a pu devenir davantage perceptible, les expériences étrangères, américaine notamment, servant d'ailleurs fréquemment d'inspirations ou de références. Deux décrets ont constitué des étapes importantes en France dans les années 1990, avec la mise en place par le décret du 22 janvier 1990 d'une série d'instances permettant de faire fonctionner un dispositif interministériel d'évaluation, que le décret du 18 novembre 1998 a ensuite tenté de simplifier (Perret, 2006).

En France, le travail d'évaluation a trouvé des soutiens dans les ministères (pour partie par l'intermédiaire des contributions des services d'inspection qui ont vu évoluer leurs fonctions dans ce sens), mais pas tous et pas de manière égale. Certains ont pu paraître précurseurs par rapport à d'autres, comme le ministère de l'Éducation nationale avec l'installation en 1987 d'une Direction de l'évaluation et de la prospective, assurant la prise en charge de travaux d'évaluation sur le système éducatif.

Trois tendances peuvent être également perçues dans les États occidentaux depuis les années 1990 : un enrichissement du travail des Cours des Comptes vers l'évaluation, un développement de l'évaluation parmi les activités des parlements et un recours également croissant à ces dispositifs aux échelons territoriaux (Lacouette-Fougère, Lascoumes, 2013). La Commission européenne a joué un rôle moteur (Perret, 2014), non seulement pour apprécier l'utilisation des fonds structurels et les programmes impulsés par ses services, mais aussi en incitant les États membres au développement des démarches évaluatives, pour partie par l'intermédiaire de guides à vocation pédagogique. Les collectivités locales sont aussi plus fréquemment devenues des commanditaires d'évaluations et certaines, au niveau régional et départemental mais dans certaines villes également, ont affiché un dynamisme en la matière (Fontaine, 1996). Au niveau international, les évaluations par les pairs (*peer reviews*) font échanger différents pays dans des logiques collaboratives permettant ainsi des expertises croisées.

Si l'on regarde les activités d'évaluation des politiques publiques, la tendance semble être à l'institutionnalisation, même si les rythmes, les formes et les facteurs explicatifs peuvent être variables (Varone, Jacob, 2004 ; Jacob, 2005) et même si cette institutionnalisation peut paraître encore fragile en France (Lacouette-Fougère, Lascoumes, 2013). Cette tendance correspond à plusieurs facteurs, qui traduisent autant d'attentes et d'intérêts, comme les souhaits de rationalisation, la recherche d'efficacité et de performance (accentuée par un contexte de resserrement des dépenses publiques), la quête de « données probantes ».

Si l'on se demande qui évalue et pour qui, il faut bien corrélativement constater que les intervenants en la matière, producteurs et potentiels consommateurs, se sont multipliés (Gregoir, 2014). Autour de l'évaluation s'est ainsi agencé et développé progressivement un système d'acteurs, qui n'exclut d'ailleurs pas la persistance de communautés concurrentes, et même si l'intérêt pour ce type de dispositifs ne s'est pas développé de manière homogène dans les institutions publiques.

C'est aussi un champ professionnel qui a gagné en structuration, dans un processus qui a permis à des acteurs publics et privés (consultants, etc.) de se spécialiser et de porter des savoirs et pratiques tendanciellement spécifiques (Matyjasik, 2013). Des prestataires interviennent ainsi sur ce qui est aussi devenu un marché. Un autre marqueur de cette dynamique, voire effort, de professionnalisation est en France la fondation en 1999 d'une Société Française de l'Évaluation (SFE), offrant un cadre de référence (avec une charte depuis 2006) pour une communauté de

consultants, agents publics, chercheurs, travaillant pour différents types d'institutions publiques selon diverses modalités contractuelles (dans le cadre d'appels d'offres ou sous des formes plus souples). La dynamique de professionnalisation se poursuit avec plus récemment l'intégration de la démarche évaluative dans les formations supérieures et même dorénavant des diplômes spécialisés de niveau Master.

## **2) Contenu des démarches évaluatives**

### 2.1) Types et visées

#### 2.1.1) Nature et spécificités de la connaissance évaluative

Les démarches évaluatives peuvent paraître voisines d'autres démarches et il n'est pas forcément facile de faire des distinctions tranchées dans un « continuum de pratiques » (Perret, 2014). L'activité est néanmoins à distinguer d'autres dispositifs qui pourraient paraître s'en rapprocher. Les promoteurs et partisans de l'évaluation tiennent couramment à la différencier des formes de contrôle qui pouvaient être familières dans le monde administratif. Elle n'est pas un contrôle de conformité et sa rationalité ne peut être réduite à une rationalité juridique, *a fortiori* si la seule préoccupation est la régularité de procédures.

L'évaluation est aussi distinguée de l'audit, dont les logiques viennent davantage du monde des entreprises. Si ces formes d'expertises professionnelles peuvent paraître proches, l'audit s'avère en fait plus axé sur le respect de normes (organisationnelles, procédurales, financières, etc.) par une structure (privée ou publique) et c'est un regard extérieur, indépendant, qui doit être la garantie de la démarche.

L'évaluation n'est pas plus censée se réduire à du contrôle de gestion. Les aspects économiques et financiers peuvent y être une dimension importante, mais ne sont généralement pas les seuls envisagés.

Appliquée à des interventions publiques finalisées, la démarche évaluative est à la fois « activité de connaissance » et « pratique d'intervention » (Duran, 2010, p. 175). Son utilité est appréhendée de façons variées. Dans certains cas, l'évaluation est une obligation à remplir suite à l'octroi de fonds publics. Les justifications les plus courantes tiennent aux souhaits de pouvoir mieux formuler ou améliorer des éléments d'intervention publique. Toutefois, derrière cet affichage, elle peut être une pièce supplémentaire dans un processus de légitimation (intervenant *a priori* ou *a posteriori*). Elle peut être aussi la base argumentaire recherchée pour optimiser ou réaffecter des dépenses publiques.

#### 2.1.2) Les moments possibles de l'évaluation

Classiquement, trois moments possibles pour l'évaluation peuvent être distingués.

L'évaluation *ex post* correspond au moment qui vient le plus spontanément à l'esprit : elle est réalisée après qu'un programme soit terminé ou à un stade suffisamment avancé d'une politique.

Des formes d'évaluation *ex ante* peuvent aussi être réalisées avant qu'une politique soit engagée ou une mesure mise en place. Il s'agit alors davantage d'étudier en amont la faisabilité et/ou de pouvoir anticiper des effets. Études et analyses d'impact sont dans cette logique et ont connu un développement notable (Lianos, Karliuk, 2014).

L'évaluation peut aussi intervenir « chemin faisant » (*in itinere*), de manière concomitante au déroulement d'une politique, par exemple à mi-parcours d'un programme si celui-ci a un calendrier bien défini. Dans cette logique, des réalisations paraissent suffisamment avancées pour justifier un bilan intermédiaire pouvant éventuellement permettre de réorienter une action ou repositionner des dispositifs, autrement dit plutôt dans un esprit de pilotage au fur et à mesure d'un processus.

S'il s'agissait d'espérer un déroulement logique et idéal, les trois moments devraient pouvoir être enchaînés l'un à la suite de l'autre (évaluation *ex ante*, puis *in itinere*, puis *ex post*). Dans les faits, rares sont les cas où ce type de schéma linéaire parvient à fonctionner, notamment parce que d'autres temporalités et considérations sont susceptibles de venir interagir.

### 2.1.3) Objets de l'évaluation et variété des critères utilisables

Effets et causes sont des axes centraux d'attention dans le travail évaluatif. Il s'agit de pouvoir apprécier des résultats et, dans la mesure du possible, de reconstituer des liens de causalité pour les expliquer. La difficulté est alors de parvenir à dégager les effets « propres » ou « purs » d'un programme ou d'une mesure. Ce qui manque est en effet le « contrefactuel », autrement dit ce qu'aurait été la situation sans ce programme ou cette mesure, ou avec d'autres options (on verra plus loin que des méthodes ont été développées pour essayer de contourner cet obstacle). Il faut aussi réussir à faire la part des facteurs extérieurs ou contextuels (économiques, techniques, sociaux, etc.), qui peuvent avoir eu une influence plus importante que la politique elle-même (typiquement une baisse du chômage permise non par des mesures ponctuelles, mais par un rebond de la croissance, elle-même favorisée par un environnement économique international plus favorable). Trouver la bonne temporalité pour caler l'analyse est une autre difficulté. Une politique jugée défavorablement à court terme peut en fait avoir besoin d'un temps long pour faire sentir ses effets. Des effets peu perceptibles au départ peuvent s'avérer structurants au fil du temps et même compliquer les choix ultérieurs. À l'inverse, des résultats obtenus rapidement peuvent s'évanouir dans la durée. Bref, le regard peut notablement changer selon les bornes temporelles choisies.

Comme chaque fois qu'il s'agit de porter une appréciation et un jugement, il faut aussi pouvoir disposer de critères, et *a fortiori* adaptés. Dans le cas de l'évaluation d'actions publiques, différents critères ont été formalisés pour apprécier leurs effets et se retrouvent couramment combinés ou articulés dans des grilles elles-mêmes plus ou moins formalisées (par exemple dans les travaux de l'Organisation de Coopération et de Développement Économiques [OCDE]).

- Le critère de **pertinence** est une manière de revenir sur les modalités de conception d'une action ou d'un dispositif : il doit permettre de vérifier que ses objectifs correspondent à des attentes ou des besoins, ou concordent avec les problèmes identifiés.

- La **cohérence** peut être appréciée de manière interne (par rapport à des principes ou des valeurs, ou entre différents types d'objectifs) ou externe (par rapport à d'autres structures ou organisations qui peuvent avoir un lien avec les interventions étudiées).

- L'**effectivité** renvoie à la possibilité de percevoir des changements liés à l'action publique évaluée et permet de commencer à aborder le degré de réalisation. Un tel critère s'avère plus pertinent s'il est rapporté à des objectifs.

- Le critère d'**efficacité** sert à déterminer dans quelle mesure des objectifs ont pu être réalisés. En rappelant la visée d'un problème initial à résoudre, placer l'attention sur l'efficacité peut par exemple changer la façon d'apprécier des Plans de Déplacements Urbains (PDU), comme le fait remarquer Jean-Marc Offner (2002, p. 121) : « Si l'on suit le PDU dans l'effectivité de sa mise en

œuvre, il n'est pas inintéressant de mesurer le nombre de kilomètres supplémentaires de pistes cyclables ; d'abord pour évaluer le caractère prescriptif du PDU compris comme programme d'action, ensuite parce que la visibilité d'une action en faveur des deux-roues n'est pas un objectif illégitime. Mais s'il s'agit d'évaluer l'efficacité du PDU, c'est l'évolution de la pratique du vélo qui doit retenir l'attention, et l'évolution concomitante des comportements automobiles ».

- Le critère d'**efficience** intègre comme dimension supplémentaire celle des moyens employés, étant entendu que ceux-ci doivent l'être le plus rationnellement possible, et avec même fréquemment l'idée de pouvoir réduire les coûts à un niveau optimal.

- La perspective s'élargit lorsqu'il s'agit de s'intéresser à l'**impact**, qui permet notamment de replacer les retombées dans une temporalité moins immédiate et de considérer aussi des effets moins prévus, voire peut-être négatifs ou indésirables.

- La **pérennité** ou la **viabilité** doit enfin permettre d'estimer dans quelle mesure les effets souhaités pourront perdurer, au-delà du court terme donc.

#### 2.1.4) Des approches en concurrence

La finalité de l'évaluation mais aussi le type de destinataires envisagés ou privilégiés peuvent induire différentes conceptions de cette démarche. Ces conceptions peuvent ainsi varier en fonction des usages escomptés.

Les efforts de conceptualisation autour de l'évaluation ont amené une distinction entre évaluation récapitulative (*summative evaluation*) et évaluation (endo)formative (*formative evaluation*), dans des termes similaires d'ailleurs à l'évaluation des apprentissages en situation scolaire. La première correspond davantage au souhait de pouvoir rendre compte d'une politique ou d'un dispositif à des acteurs qui en sont plutôt extérieurs (autorités publiques, décideurs politiques, etc.), notamment à des fins d'aide à la décision. La seconde vise à améliorer les actions entreprises en jouant sur le retour d'information en direction des acteurs impliqués.

En France, une partie des positions a aussi été structurée par des débats qui se sont déroulés dans les années 1980 et 1990, par rapports officiels interposés (rapport Deleau [1986] et rapport Viveret [1989]) et spécialement sur le rôle et la place de l'évaluation. Patrice Duran et Eric Monnier (1992) avait repéré une ligne de partage entre conceptions « gestionnaire » et « démocratique », laquelle a cependant laissé certains protagonistes suffisamment insatisfaits pour que vienne s'ajouter une autre (« pluraliste »).

Chacune a des implications en termes méthodologiques. La première conception (« gestionnaire ») paraît pour une bonne part résulter du souci principal d'allouer les moyens budgétaires de manière optimale, débouchant donc sur des pratiques pouvant rappeler davantage les audits de gestion et, en phase avec des préoccupations de rationalisation, cherchant une assise plus scientifique que politique. Dans une forme de réaction est venue une autre conception où le souci affiché portait en particulier sur la qualité du débat démocratique, qu'il s'agissait donc de pouvoir nourrir grâce à l'évaluation, sans réserver ses résultats à l'appareil d'État (ce qui pouvait paraître très ambitieux au regard des habitudes). Face aux inconvénients de ces conceptions, la conception « pluraliste » a pu être présentée comme une tentative de conciliation entre considérations techniques et politiques. L'évaluation y est appelée à être tout aussi collective que l'élaboration d'une politique et à fonctionner comme un processus d'apprentissage également collectif, tout en respectant la pluralité des systèmes de valeurs.

Certains auteurs et commentateurs vont jusqu'à parler de « générations » d'évaluation, comme s'il y avait une avancée à chacune d'entre elles (Baron, Monnier, 2003). Parmi les dernières, le souci de pouvoir élargir la gamme des intervenants, notamment du côté des bénéficiaires qu'il s'agirait de ne plus traiter simplement comme des « cibles », a contribué à ajouter des conceptions supplémentaires. Une ambition encore plus poussée a en effet été de pouvoir faire participer la diversité des « parties prenantes », nourrissant ainsi des réflexions et démarches en faveur de formes « participatives » d'évaluation. La logique recherchée est alors celle d'une co-production. La version la plus favorable à la « société civile » pourrait même être une évaluation « émancipatrice » (« *empowerment evaluation* »), qui a de fait commencé à trouver un écho aux Etats-Unis et où il s'agit de produire des capacités dans les populations intéressées. La mise en place et l'organisation de ces dernières formes d'évaluation ne sont évidemment pas simples et demandent moyens et préparations appropriées (Plottu, Plottu, 2009).

## 2.2) Méthodes et boîte à outils : une technicité croissante

La technicité de l'évaluation se perçoit par la pluralité des méthodes disponibles et une sophistication qui a continuellement avancé. La nature des informations souhaitées fait couramment retrouver la distinction classique entre celles de types quantitatif et qualitatif. Les pratiques évaluatives vont largement puiser dans les méthodes des sciences sociales et il peut en effet être tentant d'y appliquer la même séparation entre méthodes quantitatives et qualitatives. Des informations quantitatives peuvent être recherchées pour mesurer des ensembles de variables. Les traductions chiffrées de situations ou évolutions peuvent ainsi devenir combinables et manipulables, spécialement sous la forme de traitements statistiques. Les informations qualitatives restent essentiellement exprimées dans des formes discursives.

Les approches adoptées dépendent en principe des objectifs et angles de l'évaluation. En pratique, elles ne sont pas imperméables à des phénomènes de mode et aux modalités de structuration des idées dominantes (typiquement, dans la période récente, la recherche d'indicateurs chiffrés pouvant déboucher sur de la mesure de performance).

L'évaluation des politiques éducatives a pour partie été faite par des enquêtes quantitatives internationales sur les acquis des élèves (Mons, 2007). Une des plus connues, réalisée sous l'égide de l'OCDE avec une influence croissante en la matière, est l'enquête PISA (« Program for International Student Assessment » en anglais, « Programme international pour le suivi des acquis des élèves » en français), dont les résultats ont bénéficié d'une notable médiatisation. Dans la logique de ces grandes enquêtes à l'ambition comparative, les statistiques internationales sont utilisées pour chercher à repérer les facteurs influençant les performances des systèmes éducatifs, à la fois selon des angles macro (considérant les caractéristiques socio-économiques des pays et de leurs populations scolaires) et micro (propres à prendre en considération les établissements et les classes). Depuis la première version en 2000, l'enquête PISA est menée tous les trois ans, dans les pays membres de l'OCDE et d'autres pays partenaires, pour mesurer à partir d'un questionnaire les compétences des jeunes de 15 ans en compréhension de l'écrit, mathématiques et culture scientifique. Pour pouvoir tenir compte des particularités d'organisation des établissements, leurs dirigeants reçoivent eux aussi un questionnaire. Une des motivations de la comparaison est de pouvoir distinguer des « bonnes pratiques » et de les diffuser.

Les techniques qualitatives peuvent être utilisées pour apprécier le degré de satisfaction face à une mesure ou un dispositif. Elles sont en effet plus appropriées pour recueillir des opinions, des attentes, des besoins dans des groupes ou populations plus ou moins larges. Questionnaires, entretiens, observations directes, monographies font alors partie de la panoplie. Pour un dispositif particulier, il peut s'agir par exemple d'interviewer des échantillons de bénéficiaires et de non-

bénéficiaires, pour disposer d'éléments de comparaison des situations après un temps de mise en place.

Le suivi d'un panel a l'avantage de permettre un travail dans la durée. Les effets d'une mesure sont étudiés de manière longitudinale en répétant une même enquête à des échéances régulières sur un échantillon de population. Utile pour suivre le devenir de bénéficiaires, c'est une approche qui est apparue adaptée pour les politiques sociales ou les politiques d'emploi. En France, c'est depuis les années 1990 un outil du Ministère du travail, où les effets des politiques de l'emploi peuvent ainsi être observés et analysés à travers leurs bénéficiaires, par la Direction de l'animation de la recherche, des études et des statistiques (Dares) (Colin, Zamora, 2015). Ce type d'approche demande toutefois du temps et peut être coûteux. Méthodologiquement, il suppose des précautions pour garantir la représentativité de l'échantillon et son maintien sur la durée, ce qui suppose de pouvoir remplacer les personnes sortant du panel pour diverses raisons (changement de catégorie, déménagement, etc.). Ou alors, il faut prévoir au départ un échantillon suffisamment important pour qu'il puisse conserver des critères de représentativité malgré les départs et défections.

L'utilisation de certaines méthodes plutôt que d'autres dépend évidemment des informations recherchées. Les combinaisons sont donc courantes.

L'influence montante de méthodes provenant des sciences économiques est sensible, notamment sous la forme de l'analyse coût-bénéfice et du prisme qu'elle induit (Boudia, 2014). L'analyse coût-bénéfice est censée permettre d'apprécier de manière monétaire l'efficacité d'une politique ou d'un dispositif. Elle est plus fréquemment utilisée en amont de leur mise en œuvre, mais peut aussi l'être dans le cadre d'une évaluation *ex post*. Une telle analyse suppose de pouvoir quantifier et monétariser des coûts comme des bénéfices, ce qui n'est pas une tâche simple, et de les rendre relativement commensurables.

L'évaluation apparaît en fait comme « un terrain d'application pour de nombreuses branches de l'économie », ce qui fait une différence, par comparaison, avec l'investissement plus faible par la science politique, la sociologie ou d'autres disciplines des sciences sociales (Barbier, Matyjasik, 2010). Y a ainsi été également transposée l'approche expérimentale qui a pénétré les travaux de science économique et qui doit permettre de contourner l'obstacle de l'absence d'une situation contrefactuelle, celle qui aurait pu être observée sans la politique ou la mesure mise en œuvre. À la manière des expérimentations pratiquées dans le domaine médical, notamment pour les essais cliniques, la méthode des expériences « contrôlées » permet de comparer des effets éventuels dans deux groupes, l'un soumis à une mesure ou un dispositif, l'autre non (« groupe de contrôle »). C'est ce qui confère à la méthode une forte réputation de scientificité (à travers la quête de démonstration d'une relation de cause à effet) et qui lui a donné un attrait marqué dans le monde des économistes. De fait, son poids est également devenu important dans l'espace d'influence du monde nord-américain. Même si les méthodes d'évaluation aléatoire ont paru arriver tardivement en France, leur développement a été notable, à l'image de leur application à de grands programmes sociaux (L'Horty, Petit, 2011). Il reste toutefois difficile d'évacuer complètement l'inévitable part de doute sur l'écart entre les résultats d'une expérimentation et ce qui peut se passer après généralisation et application en vraie grandeur. En outre, les expériences « contrôlées » ne sont pas toujours envisageables, par exemple pour des raisons éthiques (Pourquoi réserver un dispositif avantageux à un groupe avant d'autres ?), de coût de réalisation, ou tout simplement parce que le dispositif à évaluer est déjà appliqué depuis un certain temps. Une autre option est alors celle des « quasi-expériences » ou expériences « naturelles », où il s'agit, pour s'approcher de conditions d'expérimentation, de retrouver un événement exogène ayant pu affecter un ou des groupe(s) particulier(s), mais pas d'autres (ce qui suppose bien entendu de continuer à s'assurer des possibilités de comparaison). Des techniques économétriques continuent à être développées pour cela : différence de différences (méthode consistant à revenir comparer avant et après une mesure

les éventuelles variations de situations pour des bénéficiaires par rapport à celles de non bénéficiaires), régression sur discontinuités, etc. (Givord, 2014)

Les techniques de micro-simulation sont une autre manière de tester des mesures d'action publique, en ayant d'abord fait un travail de modélisation, puis en le faisant fonctionner par l'utilisation de bases de données, elles-mêmes calibrées pour respecter les variétés de situations. C'est par exemple par un modèle de micro-simulation (MYRIADE) que la Caisse nationale des allocations familiales avait en 2003 essayé d'apprécier et d'anticiper les effets financiers et redistributifs d'une nouvelle prestation devant être mise en place l'année suivante, la prestation d'accueil du jeune enfant (PAJE) (Legendre, Lorgnet, Mahieu, Thibault, 2004 ; Thibault, 2008). L'accès à des quantités plus importantes de données et les développements des moyens informatiques en matière de calcul viennent faciliter ce genre d'approche.

Du reste, il n'est pas inutile de garder à l'esprit que, loin d'être simplement techniques, les choix de méthode ont aussi des implications politiques : ces choix ne sont pas sans hypothèses sous-jacentes (mais pas toujours explicitées) sur les manières d'envisager des situations ou phénomènes sociaux.

La dimension organisationnelle a aussi son importance dans les processus d'évaluation (Chanut, 2002). La rationalisation de ce type de pratique se voit à l'attention croissante accordée à son « design ». Les différentes étapes, du questionnement à la remise de documents finaux, en passant par la répartition des rôles pour les différents intervenants et le déroulement des investigations, sont désormais balisées (avec certes encore des variations) par moult guides méthodologiques et manuels émanant d'acteurs engagés à des degrés variables dans la sphère publique (dans cette optique presque pédagogique, voir par exemple Perret, 2014).

### **3) En aval des démarches évaluatives**

#### **3.1) Usages**

Les usages des évaluations varient évidemment en fonction des interrogations émises, des informations recherchées et des modalités d'intégration dans des processus politico-administratifs existants. Il peut s'agir d'expliquer l'apparent échec ou réussite d'une politique, mais pas seulement.

Les études d'impact ont gagné une place croissante dans l'outillage de l'action publique, bien au-delà des applications qui, initialement, concernaient davantage les projets d'infrastructures et d'aménagements. Elles sont aujourd'hui présentes de manière de plus en plus routinisée dans le répertoire instrumental des pays développés, pour examiner programmes, dispositifs ou mesures avant leur mise en œuvre.

Dans ces pays, les préoccupations relatives à la qualité de la production législative ont favorisé un intérêt grandissant pour l'analyse d'impact de la réglementation (AIR). Des éléments d'appréciation sur les effets et conséquences possibles d'une réglementation peuvent en effet être fournis par une évaluation *ex ante*. La démarche peut alors être un appui pour améliorer des choix. Elle peut aider à clarifier les résultats souhaités et les coûts correspondants, voire à mieux formuler les schémas de causalité sur lesquels sont censés jouer les leviers d'intervention envisagés. De ce point de vue, ce type d'analyse d'impact paraît alors un gage de transparence et une manière d'inciter les décideurs à assumer leurs responsabilités. Aux États-Unis, un *Executive Order* de 1978 a donné une traduction procédurale à ces préoccupations. Des organes internationaux comme l'OCDE ont également poussé dans ce sens (guides méthodologiques à l'appui), de même que, plus récemment, dans le sillage de l'idée de « *smart regulation* » promue au niveau européen (Renda, 2014). L'instrument

bénéficie d'un contexte favorable que traduit le slogan « réglementer moins, réglementer mieux ». L'étude d'impact tend ainsi à faire figure de « norme de gouvernance » (Lianos, Karliuk, 2014), même si l'instrument correspond en fait à une grande diversité de pratiques (Lianos, Fazekas, 2014). En France, la loi organique du 15 avril 2009 demande que les projets de loi fassent l'objet d'une étude d'impact sous la forme de documents joints. Quelque soient les pays, les étapes d'une analyse d'impact sont généralement similaires et laissent souvent transparaître une influence du modèle américain. La première étape revient sur le problème à traiter et les justifications de l'intervention réglementaire. La deuxième examine les options possibles. La troisième s'intéresse aux impacts de l'intervention, avec un prisme qui prend tendanciellement la forme d'une analyse coût-bénéfice. La quatrième essaye de dégager la voie optimale. Une dernière organise généralement une communication des résultats du travail pour qu'ils puissent être commentés.

Dans les pays recourant de manière croissante à cette procédure, la tendance a aussi été à accorder et aménager une place pour la consultation du public au sein des études d'impact (Denolle, 2011). En France, à l'Assemblée nationale, où cette préoccupation a été affichée, observations et contributions du public sont censées pouvoir s'insérer dans la procédure d'examen des projets de loi par les députés. Un site Internet dédié permet d'accéder aux dossiers législatifs des projets déposés et de laisser sur chacun une contribution (avec un nombre de caractère limité néanmoins).

Si les analyses coût-bénéfice ont pris une place importante dans ce genre de démarche, leurs utilisations ont montré qu'elles encourageaient le risque d'être instrumentalisées par des acteurs de l'exécutif. Aux États-Unis, les dernières présidences républicaines, sous prétexte de rationalisation, ont eu tendance à les utiliser comme une façon de freiner le travail réglementaire des agences de régulation, dans des domaines qui n'étaient pas anodins : environnement, sécurité alimentaire, santé, etc. L'outil a ainsi paru davantage servir à protéger certains intérêts économiques (Boudia, 2014). Les études d'impact ont au demeurant un degré de technicité plutôt élevé, ce qui les rend potentiellement coûteuses.

Les couplages sont aussi possibles entre plusieurs types d'évaluation. Les mesures de réduction du temps de travail (RTT) engagées en France sous le gouvernement de Lionel Jospin ont pu être appréciées à la fois sur la base de scénarios *ex ante* et d'évaluations *ex post* des effets sur l'emploi (Gubian et al., 2004).

Le souci de performance qui a de plus en plus visé l'action publique est venu conforter des usages de l'évaluation comme outil de pilotage (Baslé, 2012). Ce pilotage a eu tendance à passer conjointement par un recours accru à des batteries d'indicateurs (Brocas, von Lenep, 2010 ; Le Galès, 2016), typiquement pour réaliser le suivi d'un programme. Du fait de la complexité et de l'enchevêtrement des processus sociaux, imputer des effets à une cause ou une série de causes reste cependant souvent difficile.

### 3.2) Un répertoire instrumental extensif : convergences entre démarches évaluatives et logiques gestionnaires

L'évaluation n'est pas une pièce indépendante. On l'a dit, elle fait partie d'un agencement plus vaste et est à rapprocher d'un autre mouvement en faveur de la production et de l'usage d'indicateurs. Le passage par ces derniers marque la recherche d'informations simplifiées ou synthétiques (souvent sous formes chiffrées) sur des situations ou des tendances, pour des motifs qui peuvent aller du contrôle à la communication, en passant par l'aide à la décision. Ce déploiement d'indicateurs chiffrés reflète une extension de la statistique comme outil de gouvernement (Desrosières, 2008). Comme si tout phénomène ou fait social devait être mesurable (et, par suite, évaluable)... Dans beaucoup de pays, la mesure de la performance, appuyée sur ces

appareillages d'indicateurs, est devenue un instrument d'action publique qui porte une nouvelle vague de rationalisation, mais induit aussi une pression concurrentielle accrue (Le Galès, 2016).

En France, la loi organique relative aux lois de finances (LOLF) votée en 2001 participe de cette dynamique et l'a confortée. La réorganisation de l'architecture budgétaire en missions et programmes supposait que l'État français puisse fonctionner à partir de 2006 sur la base d'objectifs et d'indicateurs, le tout devant pouvoir être articulé et permettre de mesurer des résultats (Calmette, 2008). La démarche, tout en passant « d'une logique de moyens à une logique de résultats », visait aussi à ce que le gouvernement et l'administration puissent rendre des comptes au Parlement (Migaud, 2013). Ce faisant, le regard sur l'action publique paraît plus utilitariste (Calmette, 2006).

Le benchmarking est un autre exemple d'importation et de diffusion d'une technique managériale du monde des entreprises (chez Rank Xerox, d'abord, dans les années 1980) vers la sphère publique. La technique, s'appuyant également sur la production d'indicateurs et de tableaux de bord, consiste à comparer et étalonner des performances dans une optique d'amélioration, par l'émulation notamment. Sa diffusion a conforté l'orientation des logiques évaluatives vers une vision particulière, plutôt quantitative, de l'efficacité (Bruno, 2008 ; Bruno, Didier, 2013).

Reprenant et élargissant la logique des « audits de modernisation », qui avaient été promus de 2005 à 2007, la « révision générale des politiques publiques » (RGPP) a traduit en France l'ambition de passer en revue et remettre à plat l'ensemble des politiques étatiques, leur organisation et les moyens qu'elles mobilisent. Le déploiement du dispositif devait permettre de réaliser des économies dans les dépenses publiques et de « réformer » l'administration (en regroupant les structures, en privilégiant les tâches « stratégiques » lors de l'allocation des moyens, en réduisant les procédures administratives redondantes, etc.) (Lafarge, 2010 ; Bezes, 2010). La démarche est analogue aux « revues de programmes » (« *programmes reviews* ») déployées dans les pays qui se sont orientés vers des systèmes de gestion de la performance, comme le Canada ou le Royaume-Uni (Massey, 2010). Dans son évolution, la « Modernisation de l'action publique » (MAP), approche qui a succédé à la RGPP après l'élection de François Hollande à la Présidence de la République, a en fait retrouvé des logiques comptables dont elle était censée se détacher.

Auparavant, une brève initiative de notation des ministres avait même occupé un temps le gouvernement de François Fillon, à la fin de l'année 2007 et une partie de l'année 2008. Un cabinet privé de consultants avait été chargé de concevoir des indicateurs censés mesurer les performances des différents titulaires de portefeuille.

La conception et l'utilisation de cette masse croissante d'indicateurs ne vont pas sans soulever des difficultés, qui commencent à être bien repérées. Pour Robert Salais (2011), les politiques publiques axées sur la performance, en plus de focaliser l'attention de leurs acteurs sur l'amélioration affichable des résultats évalués par indicateurs, sont susceptibles d'être grevées par quatre problèmes : la difficulté de concilier performance et justice sociale, la difficulté d'assurer la comparabilité de données recueillies dans des situations différentes, la normativité cachée au sein des batteries d'indicateurs, et les risques de fabrication politique des chiffres comme preuves supposées. La construction d'indicateurs est évidemment la résultante de jeux de pouvoir dans lesquels, derrière les rationalités, s'expriment des visions du monde et des systèmes de valeurs (Eyraud, 2015).

### 3.3) Retombées et ambiguïtés

L'évaluation des politiques publiques, telle qu'elle s'est développée en France, semble peiner à étendre sa portée. Les habitudes gouvernementales y maintiennent encore une certaine distance vis-

à-vis de ces pratiques, même si un nombre croissant d'acteurs institutionnels affiche un intérêt pour la démarche, à défaut d'en intégrer véritablement les produits finaux. À certains échelons, même récemment, les réticences étaient encore marquées, comme chez les élus locaux par exemple, qui ont pu trouver des relais au Parlement en 2003-2004 pour amener au rejet de dispositions législatives prévoyant de généraliser l'évaluation des politiques publiques locales et de créer une instance nationale d'évaluation à cet effet (Le Lidec, 2005). Ce qu'on pourrait attendre comme un réflexe « naturel » ne l'est pas (Gaxie et Laborier, 2003), ou en tout cas pas au niveau des suppositions spontanées (et pour partie entretenues par les tendances presque performatives de certains discours officiels). Certes, le travail évaluatif paraît plus facile lorsque sont disponibles des systèmes d'information suffisamment solides pour permettre de suivre correctement les activités, ce qui n'est pas toujours le cas.

D'autres difficultés, influençant les conditions d'appropriation, sont aussi connues : la disjonction courante entre instances décisionnelles et instances d'évaluation ; le décalage des temporalités entre évaluation (qui s'élabore plutôt dans le moyen terme) et dynamique décisionnelle (tendancielle tirée vers le court terme) ; une orientation fréquemment plus rétrospective que prospective (donc potentiellement faible en recommandations) ; l'effort nécessaire pour pénétrer des rapports et rendus souvent longs, techniques, sur des sujets déjà complexes ; la difficulté à rendre capitalisables des travaux qui peuvent finir par devenir redondants (Chanut, 2002). En France, le domaine de la « politique de la ville » est symptomatique : un grand nombre d'évaluations (comparativement à d'autres secteurs) y a été conduit, mais leur poids dans les décisions politiques a été à peine discernable (Epstein, 2009).

Bref, comme le résume Bernard Perret (2014) : « Les retombées les plus importantes sont lentes, diffuses et indirectes ». L'apprentissage peut en effet prendre du temps. L'activité des évaluateurs s'avère en outre compliquée et contrainte par le rythme et l'instabilité des réformes (Bunel, L'Horty, 2011).

L'évaluation, telle qu'elle apparaît employée, n'est pas non plus exempte d'ambiguïtés. Elle n'est jamais réductible à un simple instrument et n'est pas plus neutre que n'importe quelle technique. Elle conserve inévitablement une dimension politique, puisqu'il s'agit le plus souvent de pouvoir déboucher sur des jugements, lesquels peuvent avoir des conséquences tout à fait tangibles. Un regard critique peut par exemple amener à se demander pourquoi des critères d'efficacité semblent plus systématiquement privilégiés que des critères de justice.

L'évaluation est de fait aussi une « ressource de pouvoir » (Gaxie et Laborier, 2003, p. 217). Elle peut servir à nourrir une rhétorique, dans une logique de légitimation ou de dé-légitimation notamment, quitte à y prendre ce qui arrange et à éliminer ce qui gêne. Comme le rappelle également Bernard Perret (2014) : « [...] les opinions émises sur une politique sont rarement indépendantes des positions et intérêts de leurs auteurs ».

Certains glissements sont repérables dans les usages. Des craintes ont émergé sur une utilisation trop quantitative de l'évaluation, où les préoccupations d'efficacité pourraient dériver vers une focalisation sur les réductions de dépenses, accentuant un processus auquel participait déjà la RGPP et que la MAP a tendu progressivement à rallier. Le besoin apparemment croissant de données chiffrées semble aller de pair avec un tropisme vers une évaluation plutôt gestionnaire (Ogien, 2013). De manière congruente, le raisonnement économique est devenu plus influent (Jean-Claude Barbier [2010] va jusqu'à parler d'une « banalisation [...] de l'évaluation comme l'une des formes de l'analyse économique généralisée ») et le positionnement d'un important laboratoire de recherche à Sciences Po Paris, le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques, créé en 2012), peut laisser l'impression qu'il est, dans ses orientations effectives, moins interdisciplinaire que ne l'annonce son intitulé.

En outre, ce n'est pas parce que des résultats de travaux d'évaluation sont disponibles qu'ils seront intégrés et utilisés au moment où les décisions s'élaborent. Couler l'action publique dans des programmes sous prétexte de les rendre évaluables ne garantit pas non plus qu'ils soient effectivement évalués (Epstein, 2010). D'ailleurs, l'étude des expériences passées amène à constater que les gouvernants privilégient souvent la conservation de leurs priorités. De manière symptomatique, c'est ce que montre en France le cas du revenu de solidarité active (RSA), dont la généralisation en 2009 a été décidée avant que la phase préalable d'expérimentation ne soit complètement terminée et que l'étude d'impact n'ait pu déboucher sur des conclusions utilisables (L'Horty, 2009).

Somme toute, l'évaluation, y compris l'absence d'évaluation, constitue une sorte d'analyseur institutionnel. Examiner les manières dont elle est sollicitée et employée donne aussi un éclairage sur comment opère le pouvoir politique et comment y participe la mobilisation de savoirs particuliers. On peut d'ailleurs considérer que l'évaluation peut être autre chose qu'une démonstration d'expertise. Sans être les seuls, Daniel Gaxie et Pascale Laborier (2003) ont argumenté « [pour] un usage citoyen de la sociologie de l'évaluation des actions publiques ». La question est de fait aussi celle de l'accès des citoyens aux évaluations, pas seulement en termes de mise à disposition, mais aussi de lisibilité. La mise en ligne sur Internet, pour les institutions les plus ouvertes, a pu faciliter dans une certaine mesure cet accès. En France, force est de constater néanmoins que les résultats d'évaluation sont faiblement repris par les grands médias et que les journalistes jouent très rarement un des rôles qui devrait être aussi le leur, celui d'interprètes et de médiateurs. Dommage, diraient ceux qui pensent que l'évaluation peut être aussi un instrument de participation des populations à la vie démocratique, à différentes échelles d'ailleurs.

À proximité de cette ambition démocratique, ne serait-il pas utile aussi de s'interroger sur les « trous noirs » de l'évaluation, ces mesures et dispositifs qui échappent à l'investigation évaluative alors qu'ils absorbent des masses financières loin d'être négligeables (typiquement certaines « niches fiscales ») ? En cherchant à expliquer la rareté d'évaluations rigoureuses dans de nombreux domaines de l'action publique, Lant Pritchett (2002) explique comment l'ignorance peut être « payante » ou, plus précisément, comment il peut y avoir un intérêt, pour certains acteurs porteurs d'enjeux ou de solutions particuliers, à sous-investir dans la production de connaissances empiriques fiables sur les effets de certaines interventions publiques, notamment s'il s'agit de maintenir des budgets sur les programmes privilégiés par ces acteurs.

Reste enfin une question : celle de la qualité des travaux évaluatifs réalisés. S'il s'agit de favoriser un apprentissage, il peut en effet aussi y avoir intérêt à évaluer les évaluations (Vieille question : *Quis custodiet ipsos custodes ? Who watches the Watchmen?*). Les apports de ces travaux sont-ils indiscutables ? Qui définit les normes sociales permettant de dire ce qui est bénéfique pour une collectivité ? C'est aussi l'enjeu des méta-évaluations, des évaluations de niveau supérieur en quelque sorte, puisqu'elles doivent porter sur d'autres évaluations, ce qui peut alors permettre de capitaliser les leçons tirées et d'assurer une forme d'« assurance qualité » (Baslé, 2013). Des standards ont d'ailleurs été élaborés, qui jouent non seulement comme des cadres de référence, mais tendent aussi à produire des rapprochements dans des démarches pouvant encore paraître éclatées (notamment en fonction des traditions et trajectoires historiques et institutionnelles).

Au total, on conviendra sans peine de l'intérêt d'étudier comment les démarches évaluatives peuvent éclairer les décideurs. Mais l'étude de ces démarches révèle aussi un autre intérêt : celui d'éclairer les modes d'appréhension des activités publiques et le rôle qui est attendu d'elles. Et donc d'une autre manière l'évolution de ces dernières. Comme technologies de gouvernement, les démarches évaluatives s'insèrent dans un régime où se combinent savoir et pouvoir, et il ne faudrait pas l'oublier.

## Références

Barbier Jean-Claude (2010), « Éléments pour une sociologie de l'évaluation des politiques publiques en France », *Revue française des affaires sociales*, n° 1-2, pp. 25-49.

Barbier Jean-Claude, Matyjasik Nicolas (2010), « Évaluation des politiques publiques et quantification en France : des relations ambiguës et contradictoires entre disciplines », *Revue Française de Socio-Économie*, n° 5, pp. 123-140.

Baron Gaëlle, Monnier Eric (2003), « Une approche pluraliste et participative : Coproduire l'évaluation avec la société civile », *Informations sociales*, n° 110, pp. 120-129.

Baslé Maurice (2012), *Le budget de l'État*, Paris, La Découverte, 7<sup>e</sup> éd.

Baslé Maurice (2013), « Méta évaluation des politiques publiques et qualité des évaluations », *Revue française d'administration publique*, n° 148, pp. 1017-1027.

Bezes Philippe (2010), « Morphologie de la RGPP. Une mise en perspective historique et comparative », *Revue française d'administration publique*, n° 136, pp. 775-802.

Boudia Soraya (2014), « Gouverner par les instruments économiques. La trajectoire de l'analyse coût-bénéfice dans l'action publique », in Dominique Pestre (dir.), *Le gouvernement des technosciences. Gouverner le progrès et ses dégâts depuis 1945*, Paris, La Découverte.

Brocas Anne-Marie, von Lennep Franck (2010), « Contrepoint : de l'utilité (et des limites) du recours aux indicateurs pour le pilotage des politiques publiques », *Revue française des affaires sociales*, n° 1-2, pp. 149-160.

Bruno Isabelle (2008), « La recherche scientifique au crible du *benchmarking*. Petite histoire d'une technologie de gouvernement », *Revue d'histoire moderne et contemporaine*, n° 55-4bis, pp. 28-45.

Bruno Isabelle, Didier Emmanuel (2013), *Benchmarking. L'État sous pression statistique*, Paris, La Découverte.

Bunel Matthieu, L'Horty Yannick (2011), « Pourquoi est-il si difficile d'évaluer les politiques publiques ? Le cas exemplaire des allègements de charges sur les bas salaires en France », *Reflets et perspectives de la vie économique*, Tome L, pp. 23-31.

Bureau Marie-Christine (2010), « Du travail à l'action publique : quand les dispositifs d'évaluation prennent le pouvoir », *Cahiers internationaux de sociologie*, n° 128-129, pp. 161-175.

Calmette Jean-François (2006), « La loi organique relative aux lois de finances (LOLF) : un texte, un esprit, une pratique », *Revue française d'administration publique*, n° 117, pp. 43-55.

Calmette Jean-François (2008), « La Lolf comme nouvelle approche des politiques publiques », *Informations sociales*, n° 150, pp. 22-31.

Chanut Véronique (2002), « L'évaluation : affaire d'Etat ou question d'organisation ? », *Politiques et management public*, vol. 20, n° 4, pp. 1-32.

Colin Christel, Zamora Philippe (2015), « Introduction : Les panels de bénéficiaires de politiques d'emploi à la Dares : entre continuité et innovation », *Economie et statistique*, vol. 477, n° 1, pp. 25-32.

Denolle Anne-Sophie (2011), « Les études d'impact : une révision manquée ? », *Revue française de droit constitutionnel*, n° 87, pp. 499-514.

Desrosières Alain (2008), *Gouverner par les nombres (L'argument statistique II)*, Paris, Presses de l'école des Mines.

Duran Patrice (2010a), « L'évaluation des politiques publiques : une résistible obligation. Introduction générale », *Revue française des affaires sociales*, n° 1-2, pp. 5-24.

Duran Patrice (2010b), *Penser l'action publique*, Paris, LGDJ, 2<sup>e</sup> éd.

Duran Patrice, Monnier Éric (1992), « Le développement de l'évaluation en France. Nécessités techniques et exigences politiques », *Revue française de science politique*, vol. 42, n° 2, pp. 235-262.

Epstein Renaud (2009), « À quoi sert l'évaluation ? Les leçons de la politique de la ville », *Tracés*, #09 Hors-série, pp. 187-197.

Epstein Renaud (2010), « Des politiques publiques aux programmes : l'évaluation sauvée par la LOLF ? Les enseignements de la politique de la ville », *Revue française des affaires sociales*, n° 1-2, pp. 227-250.

Eyraud Corine (2015), « Réflexions pour une sociologie de la quantification statistique et comptable », *Entreprises et histoire*, n° 79, pp. 74-87.

Flückiger Alexandre (2007), « L'évaluation législative ou comment mesurer l'efficacité des lois », *Revue européenne des sciences sociales* [En ligne], XLV-138. URL : <http://ress.revues.org/195>

Fontaine Joseph (1996), « Evaluation des politiques publiques et sciences sociales utiles. Raisons des décideurs, offres d'expertise et usages sociaux dans quelques pratiques régionales », *Politix*, vol. 9, n° 36, pp. 51-71.

Gaxie Daniel et Laborier Pascale (2003), « Des obstacles à l'évaluation des actions publiques et quelques pistes pour tenter de les surmonter », in Pierre Favre, Jack Hayward, Yves Schemeil (dir.), *Être gouverné. Etudes en l'honneur de Jean Leca*, Paris, Presses de Sciences Po.

Giddens Anthony (1994), *Les conséquences de la modernité*, Paris, L'Harmattan.

Givord Pauline (2014), « Méthodes économétriques pour l'évaluation de politiques publiques », *Economie & prévision*, n° 204-205, pp. 1-28

Gordon Ian, Lewis Janet and Young Ken, « Perspectives on Policy Analysis », *Public Administration Bulletin*, vol. 25, 1977, pp. 26-35.

Gregoir Stéphane (2014), « L'évaluation des politiques publiques : qui et comment ? », *Economie & prévision*, n° 204-205, pp. 211-224

- Gubian Alain *et al.* (2004), « Les effets de la RTT sur l'emploi : des estimations *ex ante* aux évaluations *ex post* », *Economie et statistique*, vol. 376, n° 1, pp. 25-54
- Habermas Jürgen (1978), *Raison et légitimité. Problèmes de légitimation dans le capitalisme avancé*, Paris, Payot.
- Jacob Steve (2005), *Institutionnaliser l'évaluation des politiques publiques. Étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*, Bruxelles, Presses Interuniversitaires Européennes - Peter Lang.
- Jones Charles O. (1970), *An Introduction to the Study of Public Policy*, Belmont, Wadsworth Publishing Company.
- Lacouette Fougère Clément, Lascoumes Pierre (2013), « L'évaluation : un marronnier de l'action gouvernementale ? », *Revue française d'administration publique*, n° 148, pp. 859-875.
- Lafarge François (2010), « La révision générale des politiques publiques : objet, méthodes et redevabilité », *Revue française d'administration publique*, n° 136, pp. 755-774.
- Laurent Catherine *et al.* (2009), « Pourquoi s'intéresser à la notion d' « evidence-based policy » ? », *Revue Tiers Monde*, n° 200, pp. 853-873.
- Le Galès Patrick (2016), « Performance measurement as policy instrument », *Policy Studies*, vol. 37, n° 6, pp. 508-520.
- Legendre François, Lorgnet Jean-Paul, Mahieu Ronan, Thibault Florence (2004), « Les aides publiques à la garde des jeunes enfants. Une analyse à partir du modèle de microsimulation MYRIADE », *Recherches et Prévisions*, vol. 75, n° 1, pp. 5-20.
- Le Lidec Patrick (2005), « Le pouvoir sans la responsabilité ? Le statut de l'évaluation dans la nouvelle étape de la décentralisation », *Informations sociales*, n° 121, pp. 56-64.
- L'Horty Yannick (2009), « Paradoxe de l'évaluation du Rsa », *Projet*, n° 308, pp. 21-29.
- L'Horty Yannick, Petit Pascale (2011), « Évaluation aléatoire et expérimentations sociales », *Revue française d'économie*, vol. XXVI, pp. 13-48.
- Lianos Ioannis, Fazekas Mihaly (2014), « Le patchwork de la pratique des études d'impact en Europe : proposition de taxinomie », *Revue française d'administration publique*, n° 149, pp. 29-59.
- Lianos Ioannis, Karliuk Maksim (2014), « L'émergence de l'étude d'impact comme norme de gouvernance en Europe : genèse, diffusion et acteurs », *Revue française d'administration publique*, n° 149, pp. 5-27.
- Massey Andrew (2010), « Les programme reviews au Royaume-Uni et la question de la mesure de la performance », *Revue française d'administration publique*, n° 136, pp. 837-855.
- Matyjasik Nicolas (2013), « Des évaluations de politiques publiques et des consultants. Émergence d'un champ professionnel et segmentation des rôles », *Revue française d'administration publique*, n° 148, pp. 907-921.

- Migaud Didier (2013), « Les cinq défis de l'évaluation », *Revue française d'administration publique*, n° 148, pp. 849-858.
- Mons Nathalie (2007), « L'évaluation des politiques éducatives. Apports, limites et nécessaire renouvellement des enquêtes internationales sur les acquis des élèves », *Revue internationale de politique comparée*, vol. 14, pp. 409-423
- Offner Jean-Marc (2002), « Les « suivis » des PDU : hasards et nécessités », *Métropolis*, n° 108-109, pp.116-122.
- Ogien Albert (2013), *Désacraliser le chiffre dans l'évaluation du secteur public*, Versailles, Quae éditions, 2013.
- Perret Bernard (2006), « De l'échec de la rationalisation des choix budgétaires (RCB) à la loi organique relative aux lois de finances (LOLF) », *Revue française d'administration publique*, n° 117, pp. 31-41.
- Perret Bernard (2008), « L'évaluation des politiques publiques. Entre culture du résultat et apprentissage collectif », *Esprit*, n° 350, pp. 142-159.
- Perret Bernard (2014), *L'évaluation des politiques publiques*, Paris, La Découverte.
- Plottu Béatrice, Plottu Éric (2009), « Contraintes et vertus de l'évaluation participative », *Revue française de gestion*, n° 192, pp. 31-58.
- Pritchett Lant (2002), « It pays to be ignorant: A simple political economy of rigorous program evaluation », *The Journal of Policy Reform*, vol. 5, n° 4, pp. 251-269.
- Renda Andrea (2014), « Les études d'impact des réglementations de l'Union européenne : état des lieux et pistes de réforme », *Revue française d'administration publique*, n° 149, pp. 79-103
- Rosanvallon Pierre (2006), *La contre-démocratie. La politique à l'âge de la défiance*, Paris, Seuil.
- Salais Robert (2011), « Indicateurs et conduite des politiques publiques : où sont les problèmes ? », *Informations sociales*, n° 167, pp. 60-70.
- Spenehauer Vincent (2016), « La (f)utilité gouvernementale de l'évaluation des politiques publiques, quelques leçons américaines et françaises », *LIEPP Working Paper n° 49*, Paris, Laboratoire interdisciplinaire d'évaluation des politiques publiques / Sciences Po.
- Thibault Florence (2008), « L'utilisation des modèles de micro-simulation dans l'évaluation ex ante des politiques », *Informations sociales*, n° 150, p. 68-77.
- Varone Frédéric, Jacob Steve (2004), « Institutionnalisation de l'évaluation et nouvelle gestion publique : un état des lieux comparatif », *Revue internationale de politique comparée*, vol. 11, pp. 271-292.