

## **L'expérimentation pour renouveler les politiques publiques ?**

Jean BERARD & Mathieu VALDENAIRE

**Le recours à l'expérimentation est de plus en plus souvent cité comme une forme prometteuse de réforme des politiques publiques. Elle permet de faire connaître les effets réels de ces politiques sur leurs bénéficiaires. Encore faut-il que ces résultats soient fondés sur des données objectives et réellement pris en compte.**

Les auteurs de cet essai travaillent au sein de la Mission d'animation du Fonds d'expérimentation pour la jeunesse (FEJ) (Direction de la Jeunesse, de l'Education Populaire et de la Vie Associative). Ils s'expriment ici à titre personnel et leurs propos ne sauraient engager l'administration.

L'évaluation des politiques publiques est sur toutes les lèvres et dans tous les textes qui recommandent de déterminer quelles politiques publiques sont efficaces et lesquelles ne le sont pas assez pour être poursuivies. La prolifération du terme cache une très grande hétérogénéité des pratiques : inspections, audits, tableaux de bord, palmarès, baromètres, pilotage par les résultats sont autant d'exemples d'« évaluations » qui sont d'abord des transformations des modes de gestion des entreprises et des administrations<sup>1</sup>. L'évaluation mise en œuvre dans le cadre des expérimentations et dont traite cet article vise, elle, à mesurer les effets d'une politique publique sur ses bénéficiaires.

L'idée d'« evidence-based policy », ou politique par la preuve, a rencontré aux cours des dernières années un écho de plus en plus fort. Cet intérêt s'est traduit par un développement considérable de méthodes d'évaluation<sup>2</sup> permettant d'identifier l'impact des politiques publiques. Les conditions nécessaires à la mise en œuvre de ces méthodes ne sont cependant pas toujours réunies, et leur réalisation nécessite une grande rigueur. Ainsi, dans la pratique, parmi la masse de textes et rapports d'évaluation, rares sont en réalité ceux qui permettent de mesurer des effets de manière crédible. Ceci est d'autant plus vrai lorsqu'il s'agit de politiques récemment mises en œuvre, ou en cours de déploiement.

Pour isoler l'effet d'une intervention, il est possible de procéder, de manière expérimentale et contrôlée, à son implantation progressive. Ainsi, dans le domaine scolaire, le projet STAR, mené au milieu des années 1980 dans l'État du Tennessee, a permis de tester, selon une méthodologie rigoureuse, l'effet de classes à effectif réduit sur les apprentissages et le devenir professionnel de jeunes élèves. Les résultats révélèrent un impact significativement

---

<sup>1</sup> Voir par exemple Bruno et Didier (2013).

<sup>2</sup> La première note du Conseil d'Analyse Economique (février 2013) revient sur ces méthodes, leur intérêt et leurs limites, et émet des propositions en matière d'organisation de l'évaluation des politiques publiques.

plus fort des classes de petite taille que dans la plupart des études précédentes<sup>3</sup>. Dans les années qui suivirent la publication de ces résultats, plusieurs États américains s'engagèrent dans des politiques massives de réduction de la taille des classes<sup>4</sup>. En France, la création d'un Fonds d'expérimentation pour la Jeunesse (FEJ)<sup>5</sup> suite à la nomination de Martin Hirsch comme Haut-commissaire à la Jeunesse en janvier 2009 et à l'élaboration d'un *Livre vert sur la politique de la jeunesse* a fourni un cadre d'action permettant la mise en place de telles méthodes. Le FEJ procède par appels à projets thématiques à destination des acteurs des politiques en faveur des jeunes ; il revient à ces acteurs de proposer des pistes d'action innovantes<sup>6</sup>. L'originalité principale de la démarche du FEJ tient à la place accordée à l'évaluation : celle-ci est réalisée par un organisme externe, indépendant à tous points de vue du porteur de projet et entièrement financée par le FEJ. L'ensemble des rapports d'évaluation est rendu public à l'issue des expérimentations.

Dans quelle mesure une telle méthode est-elle à même de renouveler la connaissance des effets des politiques publiques et de peser sur leur orientation ? Cet essai revient d'abord sur quelques difficultés habituellement rencontrées lorsqu'on veut isoler l'effet d'une intervention publique. Il s'interroge ensuite sur l'intérêt et la portée des résultats obtenus par le recours à l'expérimentation et les conditions dans lesquelles des résultats probants peuvent être obtenus. En effet, pour que cette démarche porte pleinement ses fruits, il faut éviter de céder à une « pensée magique » de l'expérimentation : le simple fait d'expérimenter ne suffit pas à identifier les effets de l'action testée, ni à plus forte raison à arbitrer, sans autre forme de débat scientifique et politique, entre différentes orientations pour les politiques futures dans un domaine. Une conception exigeante de l'expérimentation est nécessaire.

### **Effets incertains des politiques publiques**

La difficulté à mesurer l'impact des politiques publiques a été maintes fois exposée. Ainsi, pour reprendre cet exemple, comment identifier l'impact de la taille des classes sur l'apprentissage des élèves ? Une première estimation, reposant sur une simple comparaison, conclut à l'absence d'effet d'une telle mesure, les élèves des classes plus grandes obtenant les meilleurs résultats. Mais cet effet apparent est dû à un biais de sélection, les élèves possédant les caractéristiques les plus favorables à la réussite scolaire étant amenés, pour diverses raisons, à fréquenter des classes plus chargées. Les études faites à partir de données tenant insuffisamment compte des différences de profils entre élèves des classes les moins chargées et les plus chargées concluent ainsi à un effet positif des classes de plus grande taille. À l'inverse, les travaux parvenant à isoler l'effet propre montrent dans leur très grande majorité que la réduction de la taille des classes a un effet positif, en particulier dans les petites classes et pour les enfants issus de milieux populaires (Piketty et Valdenaire, 2006), et que cet effet est durable (Fredriksson, Öckert, Oosterbeek, 2013).

Le problème posé par ce cas précis peut être formulé d'une manière plus générale : la comparaison des trajectoires des bénéficiaires avec le reste de la population ou, au mieux, une

---

<sup>3</sup> Voir par exemple Finn et Achilles (1990) ou Krueger (1999).

<sup>4</sup> Ainsi l'Etat de Californie lance en 1996 un programme de réduction de la taille des classes qui touchera plus de 10 millions d'élèves scolarisés dans les trois premières années de l'enseignement primaire, soit plus de 90% des élèves (Unlu, 2005).

<sup>5</sup> À ce jour, 581 projets expérimentaux ont été soutenus, qui donneront lieu à 297 rapports d'évaluation. L'ensemble des rapports d'évaluation d'expérimentations terminées, dont ceux cités dans cet article, sont en ligne sur le site <http://www.experimentation.jeunes.gouv.fr/>

<sup>6</sup> 14 appels à projets ont ainsi été lancés entre 2009 et 2011. Sur le recours aux appels à projets comme vecteur d'innovation sociale, voir par exemple la note du Centre d'Analyse Stratégique (2011) consacrée aux appels à projets dans les politiques de l'emploi.

population possédant apparemment des caractéristiques comparables tient souvent lieu de preuve de son efficacité. Or les bénéficiaires d'une politique publique ne sont en général – par bonheur – pas choisis au hasard : s'ils sont orientés vers un dispositif, c'est qu'ils ont des besoins spécifiques, et donc en général des caractéristiques différentes du reste de la population, dont certaines ne sont pas observables statistiquement. Or les caractéristiques individuelles pèsent souvent (beaucoup) plus sur les trajectoires futures des personnes que n'importe quelle intervention publique. De petites différences entre les bénéficiaires et les non bénéficiaires peuvent entraîner des biais considérables dans la mesure de l'efficacité d'une politique ou d'un dispositif. L'évaluation de l'impact du redoublement sur les résultats scolaires est exemplaire de cette difficulté : si le fait de redoubler a un impact (positif ou négatif) sur la trajectoire scolaire d'un élève, il est surtout révélateur d'un niveau scolaire relativement faible, et de caractéristiques individuelles et familiales moins favorables à la performance scolaire que la moyenne des élèves. Auquel de ces deux facteurs attribuer le devenir scolaire des élèves redoublants ? Les évaluateurs d'une telle politique doivent isoler ce qui tient des caractéristiques individuelles de ce qui relève de l'intervention dont on cherche à évaluer les effets.

Si cette difficulté est acceptée en théorie, elle s'avère en pratique largement sous-estimée. Le champ de l'évaluation des politiques publiques recouvre sous un même vocable des pratiques professionnelles très diverses. Il s'agit parfois de réaliser des études qui, s'apparentant à un travail de conseil, éclairent les décideurs sur l'organisation de leurs services. Une autre forme commune d'évaluation est le suivi d'indicateurs sur les bénéficiaires des politiques publiques. De tels travaux sont fondés sur des données solides, mais ils fournissent beaucoup plus rarement une mesure crédible de l'impact d'un dispositif, en raison de la difficulté à l'isoler, parmi les différentes causes qui pèsent sur la situation des personnes.

### **L'intérêt d'expérimenter**

L'intérêt d'une démarche d'expérimentation sociale ne tient pas seulement à ce qu'elle permet d'inventer de nouveaux modes d'action, mais aussi et peut-être surtout au fait qu'elle permet de mobiliser des méthodes d'évaluation spécifiquement conçues pour tirer les leçons des projets et en mesurer les effets. La question de la nature des méthodes d'évaluation mobilisées est donc centrale.

Un avantage décisif du cadre expérimental est de permettre de comparer l'évolution d'un groupe bénéficiant d'un dispositif nouveau à celle d'un groupe disposant des ressources habituelles des politiques publiques. En procédant à une affectation aléatoire dans ses deux groupes, il est possible de constituer un groupe de bénéficiaires et un groupe témoin initialement parfaitement comparables, non seulement sur des critères observables statistiquement, mais aussi sur des dimensions qui bien qu'inobservables, pèsent sur la trajectoire des individus (par exemple, des variables de motivation, corrélées au fait d'être volontaire pour entrer dans un programme). Ainsi toute différence future constatée entre les deux groupes s'interprète sans ambiguïté comme l'effet de la mesure testée<sup>7</sup>.

Ces évaluations s'inscrivent le plus souvent dans un contexte où l'impact des politiques de droit commun est en réalité mal connu, très peu des mesures en vigueur ayant fait l'objet d'une évaluation crédible. Ainsi, l'évaluation d'un dispositif national d'aide au passage du permis de conduire (appel à projets « 10 000 permis pour réussir », lancé par le

---

<sup>7</sup> Pour une présentation plus détaillée mais non technique de ces méthodes, voir par exemple L'Horty et Petit (2011a, 2011b) ou Zamora (2011).

FEJ à l'été 2009) a permis d'identifier l'impact d'une aide financière (environ 900 euros) et d'un accompagnement fourni aux jeunes, par rapport aux aides dont ils peuvent habituellement bénéficier (les jeunes du groupe témoin bénéficient de divers dispositifs, pour un montant d'aide moyen d'environ 450 €). Ce faisant, cette évaluation fournit également des éléments sur l'efficacité des nombreux dispositifs existants<sup>8</sup>, dont l'impact n'avait jamais été mesuré. Elle montre ainsi que le dispositif augmente l'accès aux auto-écoles, la réussite au code, la réussite au permis et le fait de disposer d'un véhicule. Le taux d'obtention du permis de conduire est ainsi de 44,8 % à 24 mois, contre 29,8 % dans le groupe témoin, celui-ci étant composé de jeunes qui avaient au départ la même probabilité de réussite que les jeunes bénéficiaires. On ne constate en revanche pas d'effet notable de l'aide sur les chances d'accéder à un emploi ou sur la qualité des emplois occupés (salaire, type de contrat de travail, durée du travail et statut d'emploi) (L'Horty et al., 2012).

La collecte de données réalisée dans le cadre des expérimentations les plus ambitieuses permet de décrire des impacts inobservables avec l'appareil statistique existant. Les données disponibles pour évaluer les politiques publiques restent en effet le plus souvent d'origine administrative, et se révèlent donc fréquemment imparfaites pour décrire l'ensemble des effets des expérimentations sur ses bénéficiaires. Par exemple, les enquêtes réalisées dans le cadre de l'expérimentation d'un Revenu Contractualisé d'Autonomie (RCA)<sup>9</sup> dans les Missions locales ont pour ambition de mesurer l'impact de l'allocation attribuée aux jeunes bénéficiaires sur de multiples aspects de leur situation personnelle : situation d'emploi, efforts de recherche d'emploi et de formation, accompagnement reçu, sources de revenus, niveau et type de leurs dépenses, niveau de confiance dans diverses institutions, etc.

Enfin, approches quantitatives de mesure de l'impact et approches qualitatives peuvent être mobilisées de manière complémentaire pour évaluer les expérimentations. Ces approches donnent parfois des résultats contrastés, particulièrement intéressants parce qu'ils montrent pourquoi il est possible qu'un dispositif donne satisfaction à ses acteurs et bénéficiaires, sans qu'un impact chiffré significatif soit détectable. C'est en particulier le cas de plusieurs dispositifs qui proposent aux professionnels de bénéficier de ressources supplémentaires, ou aux bénéficiaires un transfert monétaire ou un service nouveau. Ces dispositifs recueillent le plus souvent un accueil positif de la part des professionnels comme des publics auxquels ils sont destinés. Leur effet sur la situation des bénéficiaires (en termes d'apprentissage scolaire, d'insertion sociale ou professionnelle) peut cependant être inexistant. C'est ce que révèle l'évaluation d'un dispositif de soutien après la classe destiné à des élèves de Cours Préparatoire, visant à prévenir l'échec précoce en lecture. Son évaluation qualitative montre qu'il donne satisfaction aux professeurs et aux parents, qui lui imputent, au moins en partie, les progrès réalisés par les enfants. Pourtant, une étude avec constitution aléatoire d'un groupe test et d'un groupe témoin montre que les apprentissages scolaires des enfants bénéficiaires n'ont pas progressé davantage que ceux d'autres élèves, aidés par d'autres formes de soutien proposées directement par l'école ou la commune<sup>10</sup>.

---

<sup>8</sup> 13 conseils régionaux, 33 conseils généraux et 111 villes proposent des dispositifs d'aide, sous la forme de bourses ou d'aides ciblées (L'Horty et al., 2012).

<sup>9</sup> Lancée au printemps 2011 par le FEJ, l'expérimentation du RCA vise à mesurer les effets de l'allocation d'un revenu garanti pendant une durée déterminée, dont bénéficient 5000 jeunes accueillis par 82 Missions locales, sur le parcours d'insertion sociale et professionnelle des jeunes bénéficiaires.

<sup>10</sup> Sur ce point, voir les rapports qualitatifs et quantitatifs d'évaluation du dispositif « Coup de pouce CLE », en ligne sur le site du FEJ : <http://www.experimentation.jeunes.gouv.fr/604-soutien-au-dispositif-coup-de-pouce-cle-et-evaluation.html>.

## La portée des résultats

Les expérimentations sociales fournissent donc un cadre propice pour évaluer l'impact réel des politiques publiques sur leurs bénéficiaires. Mais les résultats d'évaluation doivent également permettre de préjuger des effets de la politique expérimentée si elle était appliquée à une autre échelle et à d'autres territoires.

Au moins trois séries de critiques portent sur la validité externe des résultats d'expérimentation. Une première question est celle du caractère représentatif des individus et des territoires sur lesquels porte l'expérimentation : celle-ci concerne généralement des populations volontaires, et il n'y a en général pas de raison de supposer *a priori* que ces volontaires soient représentatifs d'une population plus large. Ensuite, les individus prenant part à l'expérimentation, acteurs comme bénéficiaires ou membres d'un groupe témoin, sont conscients de leur participation à une expérimentation, ce qui peut influencer sur leur comportement. Enfin, le changement d'échelle peut induire des effets, dits d'équilibre général, qui n'affectent pas un dispositif expérimental : par exemple, dans un contexte d'offres d'emploi rares, l'accompagnement intensif des demandeurs d'emploi peut être efficace lorsqu'il concerne un petit groupe d'individus, mais l'être moins s'il est généralisé à l'ensemble de la population.

Plus que ces problèmes eux-mêmes, c'est l'ampleur de leurs implications qui fait débat. Cette discussion est particulièrement vive dans le champ de l'économie du développement, qui constitue à ce jour l'un des domaines des sciences sociales où la démarche expérimentale a été appliquée avec le plus de systématisme. Certains auteurs (voir par exemple Deaton, 2010 ou Rodrik, 2008) considèrent ainsi que l'identification de l'impact d'un dispositif est obtenue au prix d'une portée très réduite des résultats, ceux-ci ne possédant qu'une validité locale. Cependant, l'ampleur de ces difficultés dépend grandement de la nature du dispositif testé et des effets attendus, et doit faire l'objet d'une discussion au cas par cas. Lorsque le dispositif est implanté sur des territoires suffisamment divers et que peu d'individus refusent d'en bénéficier, les bénéficiaires de l'expérimentation peuvent être représentatifs de la population à laquelle le dispositif est destiné. C'est le cas, par exemple, de l'expérimentation du RCA, ou de l'expérimentation d'une aide au passage du permis de conduire, étendues à l'ensemble du territoire et nécessitant un important soutien financier, que peu de jeunes refusent. De la même manière, la nature et l'importance des effets d'équilibre général sont extrêmement variables, selon les dispositifs expérimentés. La question de la validité externe doit ainsi être prise au sérieux, afin d'en tenir compte dans la façon de concevoir les expérimentations et de présenter les résultats, mais ces problèmes sont loin de retirer de manière absolue leur intérêt aux résultats d'évaluations effectuées dans un cadre expérimental.

Il est de ce point de vue nécessaire que les expérimentations soient conçues et interprétées en gardant en tête l'ambition de tester un mécanisme de portée générale, et non seulement de vérifier l'efficacité locale d'un dispositif. Par exemple, l'expérimentation de la « mallette des parents » menée dans l'académie de Créteil : cette expérimentation, menée en classe de sixième dans une quarantaine de collèges de l'académie de Créteil en 2008-2009, permet de juger de l'efficacité d'une politique d'implication des parents en recourant à l'organisation de réunions-débats entre les parents d'élèves et les acteurs du collège. L'expérimentation ne nous apprend pas seulement l'effet de ce dispositif particulier, mais renseigne de manière plus fondamentale sur le profit qu'il y a pour les établissements scolaires à impliquer les parents les plus éloignés de l'institution scolaire, alors que les professionnels jettent parfois un regard désabusé sur leur implication dans la scolarité de leurs

enfants. Ainsi comprise, l'expérimentation constitue un moyen potentiellement très efficace de produire une connaissance sur les effets des politiques publiques, difficilement accessible par d'autres moyens. Pour répondre à une ambition de renouvellement des politiques publiques, il est cependant nécessaire que ces enseignements soient effectivement intégrés dans les processus de décision.

### **Quelle traduction dans les politiques publiques ?**

Les résultats attendus des expérimentations sont généralement exprimés de manière binaire : généralisation de ce qui « marche » / abandon de ce qui ne « marche » pas. Des conclusions aussi tranchées ne sont possibles qu'à certaines conditions. La première, mentionnée par Esther Duflo, est que les expérimentations contrôlées mettent en évidence des écarts importants d'efficacité entre des actions qui ont des coûts comparables, et même parfois montrent que les actions les moins onéreuses ont le plus d'impact. Par exemple, dans des pays en développement, « si l'objectif est de réduire l'absentéisme à l'école, le traitement vermifuge est 20 fois plus efficace par dollar dépensé que le recrutement d'un professeur supplémentaire » (Banerjee et Duflo, 2009). La seconde condition est que cet impact ait pu être mesuré avec une certitude assez forte. Si elles sont remplies, ces deux conditions mettent le décideur public devant un choix assez aisé. Par exemple, le ministère de l'Éducation nationale, au vu des résultats de l'expérimentation de la « Malette des parents », a décidé en 2010 de l'étendre à 1 300 collèges français. Les conditions les plus favorables étaient remplies : un dispositif nouveau, normé, reproductible et peu coûteux, dont une évaluation d'impact rigoureuse avait démontré des effets significatifs (Avvisati et al., 2011).

Mais, comme le montrent les expérimentations menées en France, une telle situation est loin d'être systématique. De nombreuses évaluations ne produisent pas de résultats chiffrés, notamment dans le cas de projets de petite taille. Or, il n'est évidemment pas exclu que de petits projets donnent des résultats prometteurs. Par exemple, l'expérimentation de l'atelier pédagogique de Nanterre constitue une modalité de scolarité en dehors d'un établissement scolaire, associant des enseignants, des éducateurs et un service de psychiatrie infanto-juvénile (Gilloots, 2012). Il propose à des adolescents déscolarisés ou en rupture scolaire de reprendre pied dans la scolarité tout en tenant compte des facteurs psychologiques liés à la déscolarisation. L'évaluation montre que les jeunes expriment dans les entretiens « le sentiment d'avoir trouvé un cadre et un encadrement propices à une réappropriation des règles », « la perception d'une ambiance éducative souple par opposition au cadre scolaire rigide », et « la réelle possibilité de regagner l'estime de soi » (Boudesseul et Laffite, 2012). Les suites à donner à ce projet ne vont pas de soi. La mise en œuvre d'un tel dispositif à petite échelle a permis de comprendre son fonctionnement et pourrait permettre de l'évaluer à plus grande échelle. Une étude d'impact menée sur une durée suffisante permettrait d'estimer les effets de court terme – sur le bien-être de jeunes accueillis dans une structure adaptée – comme les éventuels effets de plus long terme, qui distingueraient un tel projet des dispositifs plus anciens pour remédier à l'échec scolaire<sup>11</sup>. Au delà des bénéfices de court terme (décider de la poursuite ou de l'extension d'un projet), l'expérimentation prend ainsi son sens dans une perspective de recherche cumulative, qui intègre les enseignements obtenus à la définition des projets à venir.

Par ailleurs, les évaluations quantitatives concluent qu'un projet a eu un impact positif si celui-ci a pu être mesuré avec une marge d'incertitude statistique raisonnable. La question du caractère statistiquement significatif (qui dépend autant de la taille de l'échantillon

---

<sup>11</sup> Voir par exemple, Palheta (2011), Millet et Thin (2012), Millet et Thin (2003).

mobilisé que de l'ampleur de l'effet attendu) n'épuise pas la question de l'ampleur des impacts constatés : à partir de quand peut-on juger qu'un impact est satisfaisant au regard des moyens engagés ? Une manière de répondre à ce type d'interrogation est de mener une analyse coûts-bénéfices, permettant de comparer une action à d'autres dispositifs, mais elle n'a rien d'évident (voir par exemple Behaghel, Crépon et Gurgand, 2009, sur l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement). Mais une évaluation d'impact solide constitue une base indispensable à toute analyse des bénéfices d'une politique.

Enfin, les résultats d'évaluation apportent fréquemment des résultats négatifs. Ceux-ci peuvent avoir des origines diverses. Certains sont liés à l'ensemble des actions publiques qui préexistent à l'expérimentation. L'innovation des projets mis en œuvre est rarement radicale, en raison de la structuration progressive, au cours des dernières décennies, des champs de la formation, de l'accompagnement et de l'insertion des jeunes les moins qualifiés, qui forment le centre de la cible des actions des expérimentations. Cette structuration a donné naissance à un ensemble de pratiques professionnelles, qui s'expriment le plus fréquemment sous les termes de partenariat, d'accompagnement et de parcours. Il arrive donc souvent que les acteurs proposent, comme forme d'action innovante, des partenariats élargis, des accompagnements renforcés et la constitution de parcours plus continus et cohérents. Il n'est évidemment pas question, pour les besoins de l'évaluation, de priver des jeunes de l'apport des politiques publiques existantes. Les enquêtes comparent donc parfois des formes d'action proches entre lesquelles il est difficile de déceler un impact.

Les résultats négatifs sont parfois liés au dispositif d'évaluation lui-même. Par exemple, un échantillon trop petit peut ne pas permettre de révéler un effet réel qui pourtant, faute d'effectifs suffisants, n'est pas distinguable d'un aléa statistique. Dans le cadre d'expérimentations bien conçues, ces résultats signifient néanmoins que l'effet attendu par les promoteurs de l'action n'a pas eu lieu. Pour cette raison, ils sont souvent mal reçus et parfois contestés par ceux qui avaient misé sur l'évaluation comme outil de légitimation de leur action. En réalité, la fréquence des résultats négatifs est une conséquence de la rigueur de l'évaluation. Des travaux qui mesurent les effets d'une action publique mettent en évidence, dans le même temps, les causalités qui lui échappent et invitent à la modestie sur les effets des politiques publiques. Le travail d'analyse de ces résultats apparaît plus difficile et moins gratifiant. C'est cependant une vertu de l'expérimentation que de laisser les « fausses bonnes idées » au stade expérimental. En outre, ces résultats désignent souvent, en creux, les facteurs décisifs sur lesquels les efforts devraient porter.

Les résultats négatifs, contrastés, incertains ou d'interprétation difficile engendrent parfois, du côté des commanditaires et, plus largement, des décideurs publics, des déceptions au regard du coût de l'évaluation et de ses contraintes pratiques. L'évaluation est décevante lorsqu'elle souligne la confusion entre les objectifs des actions des pouvoirs publics (sécuriser l'insertion professionnelle des jeunes peu diplômés, permettre à chaque jeune de décider de son orientation) et leurs effets réels. Les parcours professionnels des jeunes les plus précaires ne sont pas « sécurisés » parce que l'État met en place des actions de « sécurisation », de même les parcours scolaires de jeunes en échec ne sont pas moins subis parce que l'État promeut la notion d'« orientation active ». Des évaluations sérieuses montrent ce que produisent ces actions, mais révèlent aussi fréquemment que les faits sont difficiles à changer. Mais ces échecs, nuances et incertitudes permettent de faire entrer un peu de réalité sociale dans les processus de discussion sur l'orientation des politiques publiques. Et, lorsque des

évaluations menées dans de bonnes conditions concluent à l'effet d'une action publique, elles permettent de fonder des choix sur un niveau de connaissance très rarement disponible.

### **Favoriser l'émergence d'expérimentations probantes**

L'évaluation rigoureuse d'expérimentations sociales constitue donc un cadre privilégié pour fournir aux décideurs publics une information dont ils sont le plus souvent dépourvus : une connaissance précise et non biaisée de l'impact des actions entreprises sur leurs bénéficiaires. Mais des évaluations de qualité ne surgissent pas par hasard. Des exigences méthodologiques fortes sont nécessaires pour que les résultats soient fondés sur des bases objectives. Il est par ailleurs nécessaire d'inciter les structures (ministères, collectivités locales, associations etc.) qui proposent des actions nouvelles à se prêter à de telles évaluations (le plus simple est de cofinancer également le dispositif lui-même) et d'encourager des équipes de recherche qualifiées à candidater pour les réaliser. Cette dernière condition implique un niveau de financement suffisant pour réaliser une évaluation qui peut sembler coûteuse, mais qui représente rarement plus de quelques pourcents du coût des actions mises en œuvre, dans le cas des expérimentations, et d'un millième ou d'un dix-millième du coût total dans le cas de mesures plus générales. Il est enfin de garantir l'indépendance des évaluateurs et la crédibilité des résultats qu'elle implique, dans un contexte où de nombreuses politiques publiques sont évaluées par leurs promoteurs eux-mêmes, sur la base de protocoles plus ou moins précis.

Les résultats des évaluations doivent être largement diffusés, mis en débat, confrontés au jugement des acteurs de ces politiques comme aux résultats de recherche dans le champ desquels ils s'inscrivent. Afin que les enseignements dépassent la sphère réduite des acteurs de ces projets, la démarche d'expérimentation doit être transparente, et viser à rendre publics tout les outils et les pratiques développées dans ce cadre : la logique du financement public des expérimentations est d'engendrer des externalités positives, en permettant à d'autres de s'approprier les modes d'actions qui ont fait leur preuve.

Dans le cadre des évaluations, les échanges entre acteurs du projet et évaluateurs peuvent amener les décideurs à accorder aux résultats d'évaluation une attention dont ne bénéficient pas nécessairement les autres travaux de recherche. Ces conditions ne garantissent toutefois pas une prise en compte systématique des enseignements dans le processus de décision publique. La mise à disposition de résultats interroge ainsi plus largement la relation entre connaissance scientifique et fonctionnement politique et administratif : comment faire pour que ce que l'on sait de l'efficacité des politiques publiques devienne un critère de décision important ? Comme l'expliquait Jacques Bouveresse commentant la formule de Bourdieu selon laquelle « les obstacles à la compréhension, surtout peut-être quand il s'agit de choses sociales, se situent moins, du côté de l'entendement que du côté de la volonté », « il ne faut pas seulement vouloir savoir, il faut aussi vouloir tirer des conclusions de ce que l'on sait »<sup>12</sup>.

### **Références**

Avvisati, F., Gurgand, M., Guyon, N., Maurin, E., (2010), *Quels effets attendre d'une politique d'implication des parents dans les collèges ? Evaluation de l'impact de la Mallette des parents*, Rapport pour le Haut Commissaire à la Jeunesse

---

<sup>12</sup> *Le Monde*, 30 janvier 2002.



Banerjee, A.V, Duflo, E., (2009), « L'approche expérimentale en économie du développement », *Revue d'économie politique*, vol. 119, p. 691-726.

Behaghel, L., Crépon B., Gurgand, M., (2009), *Evaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise*, Rapport final d'évaluation.

Bernard P-Y., Michaut, C., Sanselme, F., (2011), *Espace partagé des données pour les jeunes sans solution (ESPADON-J2S)*, rapport final d'évaluation.

Boudesseul, G., Lafitte, L., (2012), *Ateliers Pédagogiques de Nanterre*, rapport final d'évaluation.

Bruno I. et Didier E. , *Benchmarking. L'Etat sous pression statistique*, Paris, Zones, 2013.

Centre d'analyse stratégique, (2011), « Les appels à projet : application aux politiques de l'emploi », Note d'analyse, n°256.

Conseil Scientifique du Fonds d'Expérimentation pour la Jeunesse (2009), *Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets*. (<http://www.experimentation.jeunes.gouv.fr/IMG/pdf/guide-pour-l-evaluation-des-experimentations.pdf>).

Conseil d'analyse économique (2013), « Evaluation des politiques publiques », Les notes du conseil d'analyse économique, n°1 (<http://www.cae.gouv.fr/IMG/pdf/CAE-note001.pdf>)

Deaton, A. (2010), « Instruments, Randomization, and Learning about Development », *Journal of Economic Literature* 48, pp. 424-455.

Duflo, E., Glennerster, R. et Kremer, M. (2008), « Using Randomization in Development Economics Research: A Toolkit », in *Handbook of Development Economics*, vol. 4, pp. 3895-3962. Amsterdam and Oxford: Elsevier, North-Holland.

Finn, J. et Achilles, C. (1990). « Answers and Questions about Class Size: A Statewide Experiment ». *American Educational Research Journal*, 27, pp. 557-577.

Fredriksson, P., Öckert B., Oosterbeek H., (2013), « Long-Term Effects of Class Size », *The Quarterly Journal of Economics*, 128 (1), pp. 249-285.

Garcia, S., Montagne, S., (2011) « Pour une sociologie critique des dispositifs d'évaluation », *Actes de la recherche en sciences sociales*, n° 189, pp. 4-15.

Gilloots M., « L'atelier pédagogique de Nanterre : une expérimentation interinstitutionnelle pour les décrocheurs scolaires », *Enfances & Psy*, 1/2012 (n° 54), pp. 102-108.

Guézennec, C. (2011). « Les “appels à projets” : application aux politiques de l'emploi », Centre d'Analyse Stratégique, Note d'analyse Travail Emploi n°256.

Inspection générale de l'éducation nationale, Inspection générale de l'administration de l'éducation nationale et de la recherche, (2005), *Sorties sans qualification. Analyse des causes, des évolutions, des solutions pour y remédier*.

Krueger, A. (1999). « Experimental Estimates of Education Production Functions ». *Quarterly Journal of Economics*, 114, pp. 497-532.

L'Horty, Y. et Petit, P. (2011), « L'évaluation aléatoire : un succès qui ne doit rien au hasard », *La vie des idées*. (<http://www.laviedesidees.fr/L-evaluation-aleatoire-un-succes.html>).

L'Horty, Y. et Petit, P. (2011) « Expérimentations sociales et évaluation aléatoires », *Revue Française d'économie*, 26, pp. 13-48.

L'Horty, Y., Duguet, E., Petit, P., Rouland, B. et Tao, Y., (2012), *Faut-il subventionner le permis de conduire des jeunes en difficulté d'insertion ?*, Evaluation de l'expérimentation « 10 000 permis pour réussir », Rapport pour le Fonds d'Expérimentation pour la Jeunesse.

Millet, M., Thin, D., (2003), « Une déscolarisation encadrée. Le traitement institutionnel du “désordre scolaire” dans les dispositifs-relais », *Actes de la recherche en sciences sociales*, n° 149, pp. 32-41.

Millet, M., Thin, D., (2012), « L'ambivalence des parents de classes populaires à l'égard des institutions de remédiation scolaire. L'exemple des dispositifs relais », *Sociétés contemporaines*, n° 86, pp. 59-83.

Palheta, U., (2011), « Le collège divise. Appartenance de classe, trajectoires scolaires et enseignement professionnel », *Sociologie*, vol. 2, pp. 363-386.

Piketty, T. et Valdenaire, M. (2006), « L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français : Estimations à partir du panel primaire 1997 et du panel secondaire 1995 », Ministère de l'Education Nationale, Les Dossiers n°173.

Rodrik, D. (2008), « The New Development Economics: We Shall Experiment, but How Shall We Learn? », Harvard Kennedy School Working Paper N° 08-055.

Unlu, F. (2005), « California Class Size Reduction Reform: New Findings from the NAEP », *mimeo* Princeton University.

Zamora, P. (2011), « La méthode d'évaluation aléatoire : apports et limites », *Tracés. Revue de sciences humaines* 11, pp. 175-186.

Publié dans [laviedesidees.fr](http://laviedesidees.fr), le 25 juin 2013

© [laviedesidees.fr](http://laviedesidees.fr)