

L'évaluation aléatoire : un succès qui ne doit rien au hasard

Yannick L'HORTY et Pascale PETIT

L'évaluation aléatoire est à la mode. Initialement surtout développée dans le champ de l'économie du développement, elle s'est maintenant imposée pour nombre de politiques publiques. Yannick L'Horty et Pascale Petit présentent l'intérêt et les limites de ce nouvel outil au service de l'évaluation des politiques sociales.

La crise alimente le développement de nouveaux besoins sociaux pour des populations toujours plus nombreuses, tout en limitant dans le même temps la capacité de l'action publique à les satisfaire. Un tel resserrement des contraintes budgétaires impose aux décideurs publics de mieux rationaliser leur intervention et, plus que jamais, d'évaluer l'impact de la dépense dans tous les domaines. Parce que l'argent public se fait rare, il devient encore plus impératif de pouvoir en mesurer les effets. Cette pression à l'évaluation est particulièrement marquée dans le champ du social où les attentes sont les plus fortes et où un très grand nombre de dispositifs co-existent et sont mis en œuvre par une pluralité d'acteurs à différents niveaux territoriaux d'intervention. C'est sur ce terrain favorable que l'on assiste en France au succès des méthodes d'évaluation aléatoire. Après avoir été mises en œuvre dans de nombreux pays, notamment aux États-Unis, au Mexique et au Canada, ces méthodes se diffusent aujourd'hui en France où elles sont désormais appliquées à des centaines de programmes sociaux locaux.

Choc d'offre et choc de demande publique

Le principe de l'évaluation aléatoire est extrêmement simple. Au sein d'une liste de bénéficiaires potentiels à un programme social, on constitue deux groupes en effectuant un

tirage au sort. Le groupe-test a accès au programme, le groupe-témoin n'y participe pas. Il n'y plus qu'à comparer les deux groupes pour évaluer les effets du programme à différents horizons. Ce principe permet de résoudre bien des problèmes techniques posés par l'évaluation des politiques publiques (voir *infra*).

Cela implique néanmoins une rupture temporaire du principe d'égalité et c'est d'ailleurs pour lui donner un fondement juridique que la réforme constitutionnelle de 2003, concernant la décentralisation, a été adoptée la même année que la loi organique relative à l'expérimentation par les collectivités territoriales. Les expérimentations sociales deviennent possibles dès lors qu'elles ont un objet circonscrit dans l'espace, pour une durée limitée dans le temps et si elles sont menées en vue d'une généralisation. De plus, elles doivent s'effectuer à l'initiative d'une collectivité locale et sont nécessairement évaluées. L'expérimentation du revenu de solidarité active (RSA) prévue dans la loi du 21 août 2007 en faveur du travail, de l'emploi et du pouvoir d'achat (« loi Tèpa ») a constitué la première expérimentation sociale de grande envergure en France, même si cette expérimentation n'a finalement pas été évaluée selon une méthode purement aléatoire, dans la mesure où ni les bénéficiaires du RSA expérimental, ni les territoires-test de l'expérimentation n'ont été tirés au sort.

L'impulsion de Martin Hirsch a joué un rôle décisif dans le développement des expérimentations. Devenu en juin 2007 Haut Commissaire aux Solidarités actives contre la Pauvreté et promu en janvier 2009 Haut Commissaire à la Jeunesse, l'initiateur du RSA a soutenu de façon constante le développement des expérimentations sociales et de leur évaluation à l'aide de méthodes aléatoires. Il lance un premier appel à projet d'expérimentation sociale en 2007, doté d'un budget de six millions d'euros. Ce coup d'essai est suivi en 2009 d'une série d'appels à projets lancés par le fonds d'expérimentations pour la jeunesse, créé par l'article 25 de la loi généralisant le RSA du 1er décembre 2008, pour un budget total de 150 millions d'euros, issus pour les deux tiers du public et pour le solde de fonds privés. Près de 450 projets innovants sont ainsi financés, dont une part importante prévoit une évaluation aléatoire.

Au-delà du choc de demande publique, le développement des méthodes d'évaluation aléatoire s'explique aussi par un autre choc positif, cette fois-ci du côté de l'offre d'évaluation. La diffusion des travaux d'Esther Duflo a joué ici un rôle crucial en suscitant l'engouement de nombreux chercheurs français. L'un des messages essentiels de la

professeure au MIT, qui est l'une des fondatrices du réseau international du J-PAL, est que l'évaluation expérimentale a fait ses preuves pour analyser les causes de la pauvreté dans les pays pauvres, et qu'il convient désormais de l'utiliser pour le même objectif dans les pays riches et notamment en France. Le message sera développé dans sa leçon inaugurale de la chaire internationale « Savoirs contre pauvreté » du Collège de France (Duflo, 2009) et dans ses deux ouvrages destinés à un public large (Duflo 2010a et 2010b). Il sera relayé par de nombreux économistes français, notamment François Bourguignon, directeur de l'École d'économie de Paris et président du comité national d'évaluation des expérimentations du RSA, et Marc Gurgand, qui préside quant à lui le conseil scientifique du Fonds d'expérimentation pour la jeunesse. En outre, de nombreux colloques ont popularisé l'apport des méthodes expérimentales, notamment les rencontres de l'insertion lancées à Grenoble en novembre 2007, le colloque « Expérimentations pour les politiques publiques de l'emploi et de la formation », organisé par la Dares en mai 2008, ou la conférence nationale de l'expérimentation sociale, organisée par les deux Hauts Commissariats (solidarités actives et jeunesse) en mars 2010.

C'est bien la conjonction d'un changement du cadre institutionnel, d'un choc de demande et d'un choc d'offre positif qui permet d'interpréter le développement de l'évaluation aléatoire en France. Pour que cette conjonction fonctionne, il fallait en outre que l'appariement entre l'offre et la demande ait effectivement lieu. Or les nouvelles technologies d'évaluation ont une caractéristique particulière qui leur permet de répondre efficacement à la nouvelle demande d'évaluation. Elles peuvent s'appliquer dans un contexte où les données sont rares, voire inexistantes, et où elles sont construites par l'évaluation. Alors que l'économétrie de l'évaluation quasi-expérimentale nécessite de très grandes bases de données pré-existantes, dans lesquelles le nombre d'observations se compte en dizaines de milliers, les méthodes d'évaluation aléatoire permettent de détecter les effets d'un programme dont les bénéficiaires ne se comptent qu'en centaines. Cela correspond précisément aux contraintes des micro-programmes locaux pour lesquels il y a une demande en forte expansion dans le contexte français de décentralisation et de territorialisation croissante de l'action publique.

Les qualités intrinsèques de ces méthodes

En amont de ces éléments, ce sont les qualités intrinsèques des méthodes d'évaluation aléatoire qui expliquent leur succès. Dès lors que la question est bien de chiffrer l'impact *ceteris paribus* d'un programme social sur un ensemble de variables d'intérêt, les méthodes

aléatoires sont en mesure de satisfaire pleinement les attentes des évaluateurs. Le tirage au sort fournit en effet une excellente réponse à ce que James Heckman (1992) qualifie de « problème de l'évaluation ». Si l'on souhaite mesurer les effets d'un programme, il importe de pouvoir observer un même individu dans deux situations différentes, celle où il bénéficie du programme et celle où il n'en bénéficie pas. Mais comment effectuer ces observations, alors que pour chaque individu un seul état du monde s'est réalisé ? Comment savoir quel aurait été le comportement du bénéficiaire s'il n'avait pas eu accès au programme ? Répondre à ces questions revient à pouvoir construire un groupe de contrôle, appelé également « contrefactuel », qui imite au mieux les comportements des membres du groupe de traitement. Tout le problème de l'évaluation est de construire un groupe témoin de qualité. Dans l'évaluation aléatoire, le contrefactuel a un statut particulier. Il n'est pas donné de façon fortuite par la nature comme dans le cas des expériences naturelles. Il est construit *ex nihilo* par le chercheur. Il est observé et non inventé.

Un contrefactuel de bonne qualité permet de s'affranchir du risque de biais de sélection. La difficulté est de contrôler l'hétérogénéité à la fois sur les variables observables et sur les inobservables. Pour y parvenir en dehors de tout cadre expérimental, il est nécessaire de mobiliser des techniques de traitement des données très sophistiquées, telle que la méthode d'appariement proposée par D. Rubin, qui requiert un grand nombre d'observations. Le recours au cadre expérimental d'une évaluation aléatoire garantit quant à lui que les personnes des deux groupes auront en moyenne les mêmes caractéristiques observables et non observables. S'ils ont une taille suffisante (de l'ordre de plusieurs centaines d'individus), les deux groupes auront exactement la même structure par âge, sexe, niveau de diplôme, et aussi selon d'autres caractéristiques que l'on ne peut pas observer mais qui pourront avoir un effet sur les variables d'intérêt, par exemple la motivation, les préférences, les capacités cognitives, etc. L'intérêt majeur d'une évaluation aléatoire est de s'affranchir de ces biais de sélection, de façon à produire un chiffrage robuste avec une grande économie de moyens statistiques ou économétriques.

Au-delà de cet avantage méthodologique, le plus souvent mis en avant, les méthodes d'évaluation aléatoire modifient de façon positive et constructive le métier du chercheur qui les applique. Dans ses ouvrages et ses articles, Esther Duflo met ainsi en avant un avantage inattendu de ces approches. Puisqu'elles impliquent un partenariat étroit entre un expérimentateur, qui porte une innovation de politique publique, et un évaluateur, qui en

mesure les effets, et que ce partenariat s'établit dans la durée, avant, pendant et après la mise en œuvre de la politique, il en résulte une relation particulière entre les deux parties prenantes. L'expérimentateur intègre au fur et à mesure les pré-requis de l'évaluation tandis que l'évaluateur devient progressivement co-constructeur de la politique qu'il évalue. Le chercheur gagne une connaissance intime de la politique mise en œuvre qui lui permet de devenir encore plus créatif dans son questionnement. La qualité de sa recherche s'en trouve améliorée.

On peut ajouter que le chercheur qui pratique l'évaluation aléatoire élargit de façon significative son emprise sur son objet et ses méthodes de recherche. Non seulement il peut devenir co-constructeur de la politique publique, mais aussi il n'est plus passif vis-à-vis des données, qu'il construit ou reconstruit intégralement selon les cas. L'économie appliquée devient une recherche de terrain. Le chercheur est moins présent dans son bureau à travailler sur des modèles ou sur des programmes, et est davantage *in situ* à mettre en œuvre des protocoles expérimentaux. Du fait de cette implication accrue du chercheur, l'évaluation expérimentale devient plus transparente pour l'expérimentateur et pour le financeur. Elle est plus lisible pour le *policy maker* qui va souvent lui accorder davantage de crédibilité que les approches classiques qui ne partagent pas les mêmes propriétés.

Toute médaille a son revers

Dans notre introduction aux méthodes d'évaluation aléatoire (L'Horty et Petit, 2010), nous indiquons que chacun des avantages prêtés à ces méthodes a une contrepartie négative. Tout d'abord, la randomisation peut poser un problème d'acceptabilité et un problème éthique, qui ne se recoupent pas complètement. Un protocole d'évaluation aléatoire implique de priver une partie de la population des ressources qui pourraient lui être nécessaires pour améliorer sa situation. Or, si ces ressources ont effectivement un effet sur la trajectoire des personnes, leur privation pour le groupe témoin peut aller à l'encontre de l'objet même de l'institution expérimentatrice. La finalité première de ce type d'institution est d'améliorer le bien-être de personnes en difficulté d'insertion sociale et/ou économique. L'acceptabilité est la question soulevée par les travailleurs sociaux qui sont en contact direct avec les personnes ou encore par les élus qui souhaitent naturellement que le plus grand nombre de bénéficiaires ait un accès immédiat à l'innovation sociale.

De plus, lorsqu'une expérimentation est conduite sur un territoire spécifique par un expérimentateur spécifique, il est possible qu'elle produise des effets spécifiques, qui ne sont pas généralisables. En particulier, puisque les expérimentateurs porteurs du projet sont engagés dans sa conception et sont volontaires pour sa réalisation, ils seront favorables à ce type de protocole et prêts à fournir les efforts nécessaires à sa mise en œuvre. Leur motivation à la réussite de l'expérimentation peut les conduire à introduire plus ou moins consciemment des mesures d'accompagnement dans le groupe test, qui ne sont pas prévues dans le protocole. Les effets de la mesure testée sur le groupe test peuvent s'en trouver influencés.

Si l'évaluation aléatoire permet effectivement de s'affranchir des biais de sélection, elle expose ainsi à d'autres biais spécifiques. Puisque les expérimentations sont circonscrites à la fois dans l'espace et dans le temps, la généralisation du programme implique un changement d'échelle temporelle ou spatiale qui peut produire un changement dans les effets du programme. On parle ici d'« effets d'équilibre ». L'extension d'un dispositif local a des effets agrégés qui modifient les équilibres de marché et les prix. Au travers de ces effets, le traitement a un impact sur les non traités qui n'est pas considéré par une évaluation aléatoire. Le problème est donc de savoir si le résultat observé au niveau local sera toujours le même quand l'expérience est généralisée et que des effets agrégés entrent en compte. Rodrick (2008) soutient par exemple que la fourniture gratuite de moustiquaires anti-malaria, même si elle s'avère efficace dans le cas d'une expérimentation aléatoire, ne peut pas être étendue à l'échelle d'un pays parce qu'elle ferait disparaître les réseaux marchands de distributions qui fournissent des zones rurales.

Une autre catégorie d'obstacles à la mise en œuvre d'expérimentations et à leur évaluation selon des protocoles exigeants peut être évoquée. Il s'agit des difficultés matérielles rencontrées par les opérateurs de ces programmes et leurs évaluateurs. Parce qu'une évaluation aléatoire suppose de suivre une expérimentation avant, pendant et après sa réalisation, tout en construisant des données originales, elle implique des coûts logistiques souvent très importants et des délais de production parfois conséquents. C'est pourquoi le temps de l'évaluation n'est pas toujours conforme au temps de la décision publique. De surcroît, l'évaluateur va être nécessairement confronté à un grand nombre d'imprévus et d'impondérables. S'il pense évaluer une expérimentation, en réalité il expérimente une évaluation.

Un outil de plus dans la panoplie des évaluateurs

Finalement, il est clair que le développement d'évaluations aléatoires est un progrès indéniable qui élargit la boîte à outils des chercheurs s'intéressant aux questions sociales, mais ce n'est pas un nouveau *Gold Standard* méthodologique qui remplacerait toutes les autres approches. Les méthodes expérimentales peuvent occuper une place de choix dans l'ensemble des méthodes d'évaluation à condition de bien connaître leur portée et leurs limites pour circonscrire cette place et les utiliser à bon escient.

Toutes les politiques publiques ne relèvent pas d'une évaluation aléatoire. Les politiques macroéconomiques, monétaires ou budgétaires, de même que les grandes actions structurelles dans le domaine de la fiscalité, de la protection sociale ou encore dans celui des politiques industrielles ou commerciales, ne sont pas évaluables par ce type de méthode. Ce n'est pas parce qu'elles ne sont pas évaluables que l'on ne doit pas les mettre en œuvre. Dans la panoplie des outils d'évaluation, il y a une place à prendre pour les méthodes aléatoires. Ces méthodes ne doivent prendre toute la place, mais elles doivent prendre leur place.

Que l'on ne s'y trompe pas, porter un regard nuancé sur les apports et limites de ces méthodes ne doit pas inciter à l'inaction. Alors que les méthodes expérimentales sont utilisées depuis longtemps dans les sciences dures, en médecine ou en agronomie, et même en marketing, et qu'elles sont utilisées depuis les années soixante pour évaluer de grands programmes sociaux dans des pays riches, en Amérique du Nord principalement, leur introduction a été très tardive en France où un retard certain reste à rattraper. Or l'évaluation aléatoire reste dans bien des cas la seule approche quantitative adaptée à un programme social local dont le public cible est de quelques centaines de bénéficiaires. Ne pas mettre en œuvre ce type de méthodologie équivaut à renoncer à toute évaluation chiffrée des effets des politiques sociales, ce qui est sans doute la pire des solutions du point de vue de la décision publique.

Références bibliographiques

DEATON A., 2009, "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic Development", *mimeo Princeton University*, janvier.

DUFLO E., 2009, « Expériences, science et lutte contre la pauvreté », Leçon inaugurale au Collège de France, Chaire internationale « Savoirs contre pauvreté », 8 janvier.

DUFLO E., 2010a, *Le développement humain. Lutter contre la pauvreté (I)*, Le Seuil, « République des idées », Paris, 104 p.

DUFLO E., 2010b, *La politique de l'autonomie. Lutter contre la pauvreté (II)*, Le Seuil, « République des idées », Paris, 104 p.

HECKMAN J., 1992, "Randomization and Social Policy Evaluation," in Ch. Manski and I. Garfinkel (eds), *Evaluating Welfare and Training Programs*, Cambridge, MA: Harvard University Press.

L'HORTY Y., et PETIT P., 2010, « Évaluation aléatoire et expérimentations sociales », *Revue Française d'Économie*, à paraître.

RODRIG D., 2008, "The New Development Economics: We Shall Experiment, but How Shall We Learn?", *HKS Faculty Research Working Paper Series*, n° 08-055

Article publié dans lavedesidees.fr le 11 mai 2011.

© lavedesidees.fr