

L'expérimentation sociale : étapes et méthodes d'évaluation

Qu'entend-on par expérimenter ? Selon la définition du Larousse, une expérimentation désigne « une méthode scientifique reposant sur l'expérience et l'observation contrôlée pour vérifier des hypothèses ». D'un point de vue scientifique, il s'agit de construire un contrefactuel permettant la mesure de relation causale, c'est-à-dire de comparer une population qui va bénéficier d'un dispositif et une population qui n'en bénéficiera pas. L'appropriation du terme par le champ politique a donné lieu à une définition plus large : la mise en œuvre à petite échelle d'un dispositif et l'occasion de procéder à son évaluation. L'objectif d'une expérimentation est de mesurer les effets d'un dispositif et ses conditions de mise en œuvre afin d'envisager de le généraliser dans son ensemble ou en partie. Pour ce faire, différentes méthodes issues de plusieurs disciplines scientifiques (essentiellement l'économie et la sociologie) sont mobilisées.

DES QUESTIONS À L'ÉPREUVE DU TERRAIN

À l'origine de chaque expérimentation, il y a une question, une hypothèse formulée théoriquement qui vise à être vérifiée ou réfutée empiriquement.

La réduction du nombre d'élèves par classe dans les zones urbaines les plus pauvres améliore-t-elle les résultats scolaires des élèves (1) ? Un traitement vermifuge en direction des élèves de l'école primaire au Kenya permet-il de réduire l'absentéisme scolaire (2) ? La mise en place d'une allocation chômage dégressive, qui ne disparaît qu'une fois atteint l'équivalent d'un SMIC à plein-temps, permet-elle le retour à l'emploi des chômeurs (3) ?

D'autres types de questionnements, plus ouverts, peuvent être à l'origine d'une expérimentation dans son acception plus large : comment associer les habitants à la gestion de services au sein des quartiers prioritaires des politiques de la ville (mesures « Habitat et vie sociale », 1977) ? Comment éviter les ruptures de parcours de jeunes sortant des dispositifs de l'aide sociale à l'enfance (expérimentation « ROAD 94 » lancée par le Fonds d'expérimentation pour la jeunesse (FEJ) en 2009) ? Il s'agit davantage ici de repérer des dispositifs, des pratiques permettant de résoudre un problème.

Top down et bottom up : deux logiques

Les deux types de questions correspondent à des logiques différentes. La première est plutôt descendante : il s'agit d'expérimenter une politique et d'en mesurer l'impact afin de la généraliser. La seconde est ascendante : elle consiste à faire remonter des actions du terrain afin de résoudre un problème pour en dégager des enseignements et éventuellement les diffuser voire les développer. Si ces logiques peuvent être indépendantes, elles peuvent aussi être successives.

Ensuite, il peut être nécessaire de faire remonter puis de tester des idées de dispositifs et d'actions mis en œuvre de manière temporaire, afin de les évaluer. Ainsi, la mesure des effets et l'analyse des conditions de mise en œuvre sont les préalables indispensables à la dernière étape, qui est la mesure d'impact scientifique. En effet, la mise en place d'un contrefactuel ne peut être effectuée sans un repérage des porteurs capables de mettre en œuvre une expérimentation de manière fiable, ni sans délimitation précise d'une problématique et définition d'indicateurs pertinents au regard du sujet traité.

DES MÉTHODES D'ÉVALUATION PLURIELLES

L'expérimentation aléatoire ou quasi aléatoire : la mesure de l'efficacité

L'expérimentation aléatoire ou randomisée permet de se prononcer sur l'efficacité, l'impact et les variables d'intérêt d'un dispositif. Cette définition de l'expérimentation sociale, telle qu'elle a été développée d'abord aux États-Unis, s'inspire des essais cliniques dans le domaine médical. Plusieurs dénominations sont utilisées dans la littérature : « expérimentation aléatoire » (Jatteau, 2013), « évaluation expérimentale » (Fougère, 2010) ou encore « évaluation aléatoire » (L'Horty, Petit, 2011).

Cette méthode repose sur l'observation d'un groupe test bénéficiant du dispositif et d'un groupe témoin qui n'en bénéficie pas. L'expérimentation aléatoire consiste à produire un contrefactuel permettant de comparer les effets d'un dispositif à « ce qui serait advenu dans un état de nature » (Gurgand, 2014), sans le dispositif.

L'expérimentation aléatoire est donc plus qu'une démarche d'évaluation, puisqu'elle fait partie intégrante de l'élaboration du dispositif en lui-même.

Le Fonds d'expérimentation pour la jeunesse (FEJ), créé en 2009, est considéré comme un « *public sector innovation lab* » (McGann *et al.*, 2018 ; Fuller *et al.*, 2016). Il s'organise et se positionne selon deux logiques : une logique descendante quand des expérimentations sont lancées par le gouvernement pour tester une politique avant de la généraliser, et une logique ascendante lorsqu'il s'agit par le biais d'acteurs de terrain de faire remonter des dispositifs grâce à des appels à projets. Un ou plusieurs ministères, des acteurs institutionnels, du monde associatif, des entreprises privées mais aussi des équipes d'universitaire ou de cabinet collaborent à la mise en œuvre d'une expérimentation.

Pour cette raison, elle n'est pas toujours réalisable, lorsque le projet a démarré avant l'évaluation par exemple, ou si la population non bénéficiaire risque de profiter indirectement du dispositif en côtoyant la population bénéficiaire, ou encore si, pour des raisons éthiques, d'autres méthodes sont préférées afin de ne pas priver certaines personnes de droits déterminants. Dans ces cas-là, des méthodes dites « quasi aléatoires » peuvent être utilisées afin de reconstituer des groupes comparables *a posteriori*. Les méthodes le plus fréquemment mises en œuvre dans ce cadre sont les suivantes :

La double différence vise à observer deux groupes, l'un bénéficiaire d'un dispositif et l'autre témoin à deux moments : avant et après le dispositif. En supposant que les deux groupes auraient évolué de la même manière sans le dispositif, on mesure la différence entre les valeurs observées (Behaghel *et al.*, 2013).

La méthode d'appariement ou *matching* consiste à trouver pour chaque individu testé un individu identique du point de vue des caractéristiques observables (âge, sexe, catégorie socioprofessionnelle, etc.) afin de faire une moyenne des effets individuels du dispositif pour obtenir un effet global moyen du dispositif expérimenté (Arceneaux *et al.*, 2006).

La régression par discontinuité est utilisée dans le cadre de dispositifs pour lesquels un seuil ou une caractéristique (âge, lieu d'habitation, etc.) définit les personnes éligibles. L'effet du dispositif est observé par la comparaison entre la population test et des individus non bénéficiaires, mais qui ont des caractéristiques très proches du seuil d'éligibilité (en cela on considère qu'ils sont comparables aux individus testés).

Les méthodes d'expérimentations quasi aléatoires, plus encore que les méthodes aléatoires, nécessitent un nombre important de bénéficiaires et un budget conséquent. De plus, l'impact visé par une expérimentation n'est parfois visible que sur le long terme, ce qui implique des suivis de cohortes sur plusieurs années, voire décennies.

Méthode de comparaison avant et après : regarder ce qui bouge dans le système

La comparaison d'une situation avant et après la mise en place du dispositif permet de regarder les effets produits sur le système et l'ensemble de ses membres par la mise en perspective des expériences des uns et des autres. L'objectif est de comparer la situation à deux moments donnés en regardant les positions et les points de vue des différents membres du système : évolution des pratiques des acteurs, etc. Le fait que des éléments extérieurs au dispositif puissent impacter les acteurs conduit à prendre des précautions pour parler d'« impact » sauf si une population témoin est constituée. En revanche, cette méthode permet d'observer les différents effets à court et moyen terme, y compris ceux qui n'avaient pas été prévus initialement.

Méthode *in itinere* : comprendre les conditions de mise en œuvre

L'évaluation *in itinere* porte davantage sur la cohérence, la pertinence d'un dispositif et permet de produire de la connaissance sur une situation et sur un dispositif en cours de construction, voire en co-construction. Les termes d'évaluation « chemin faisant », « concomitante », « en continu » ou « au fil de l'eau » sont aussi utilisés pour désigner ce type d'évaluation. Il s'agit d'observer le dispositif et sa mise en œuvre du point de vue de l'ensemble des acteurs inscrits dans un contexte spécifique (territoire). Ce type de méthode peut conduire à repérer des besoins ou des atouts invisibles au départ, voire de faire rectifier une question théorique en lien avec les problèmes du terrain, elle peut également permettre de faire émerger une problématique de terrain.

La réalisation d'une évaluation implique un certain nombre d'outils de recueil de données tels que le traitement de données administratives, le questionnaire, l'entretien individuel et collectif et l'observation. Ces méthodes, et leurs outils, peuvent être mises en place indépendamment ou simultanément, afin d'évaluer l'impact strict d'un dispositif, mais aussi de comprendre les effets attendus ou inattendus et les conditions de mise en œuvre indispensables à toute réflexion sur les politiques publiques.

La combinaison des méthodes permettant de mesurer les différents effets d'un dispositif ainsi que ses conditions de mise en œuvre est nécessaire à l'appréhension la plus complète de la réalité, ce afin de produire une connaissance utile aux décideurs politiques et aux acteurs de terrains.

Aude Kerivel et Samuel James,
chargés d'études et d'évaluation, INJEP

(1) Expérimentation Tennessee Student/Teacher Achievement Ratio, menée à partir de l'année scolaire 1985-1986 aux États-Unis.

(2) Expérimentation mise en œuvre par le The World Health Organization et le Abdul Latif Jameel Poverty Action Lab, en partenariat avec le ministère de la santé Kenyan.

(3) Expérimentation du revenu de solidarité active (RSA) en France par le haut-commissariat aux solidarités actives et contre la pauvreté.

POUR ALLER PLUS LOIN

- Arceneaux K., Gerber A. S., Green D. P., 2006, « Comparing experimental and matching methods using a large-scale voter mobilization experiment », *Political Analysis*, n° 1, vol. 14, p. 37-62.
- Behaghel L., Crepon B., Gurgand M. *et al.*, 2013, « L'accompagnement personnalisé des demandeurs d'emploi », *Revue française d'économie*, vol. 28, n° 1, p. 123-158 [len.lignel](#).
- Fougère D., 2010, « Les méthodes économétriques d'évaluation », *Revue française des affaires sociales*, n° 1, p. 105-128 [len.lignel](#).
- Fuller, M., Lochard, A., 2016, *Public policy labs in European Union member states*, Publications de l'Union européenne, Luxembourg [len.lignel](#).
- Gurgand M., 2014, « Que nous apprennent les expérimentations sociales ? », Cours au collège de France, 21 février [len.lignel](#).
- Jatteau A., 2013, *Les expérimentations aléatoires en économie*, Paris, La Découverte.
- Kerivel A., 2018, *Guide méthodologique relatif aux évaluations du Fonds d'expérimentation pour la jeunesse (FEJ)* [len.lignel](#).
- L'Horty Y., Petit P., 2011, « Évaluation aléatoire et expérimentations sociales », *Revue française d'économie*, n° 1, vol. 26, p. 13-48 [len.lignel](#).
- McGann M., Blomkamp E., Lewis M. J., 2018, « The rise of public sector innovation labs : experiments in design thinking for policy », *Policy Science*, n° 3, vol. 51, p. 249-267.