

TRAVAIL
ET EMPLOI

Travail et Emploi

135 | juillet-septembre 2013

Expérimentations de terrain et politiques publiques du travail et de l'emploi. Apports récents et mises en perspective

Expérimentations de terrain et politiques publiques du travail et de l'emploi. Apports récents et mises en perspective

Introduction

Field experiments and public policies in work and employment. Recent contributions and perspectives. Introduction

Agnès Labrousse et Philippe Zamora



Édition électronique

URL : <http://journals.openedition.org/travailemloi/6048>

DOI : 10.4000/travailemloi.6048

ISSN : 1775-416X

Éditeur

DARES - Ministère du Travail

Édition imprimée

Date de publication : 15 septembre 2013

Pagination : 5-13

ISSN : 0224-4365

Référence électronique

Agnès Labrousse et Philippe Zamora, « Expérimentations de terrain et politiques publiques du travail et de l'emploi. Apports récents et mises en perspective », *Travail et Emploi* [En ligne], 135 | juillet-septembre 2013, mis en ligne le 01 juillet 2013, consulté le 03 mai 2019. URL : <http://journals.openedition.org/travailemloi/6048> ; DOI : 10.4000/travailemloi.6048

Expérimentations de terrain et politiques publiques du travail et de l'emploi. Apports récents et mises en perspective

Introduction (*)

Agnès Labrousse (**), Philippe Zamora (***)

La création d'un Fonds d'expérimentation pour la jeunesse (FEJ) en 2008 témoigne de l'essor récent des expérimentations de terrain en France (L'HORTY, PETIT, 2010). Pour la première fois, plusieurs expérimentations par assignation aléatoire y ont été conduites, stimulées par l'écho des travaux du J-PAL (*Abdul Latif Jameel Poverty Action Lab*) sur la pauvreté (BANERJEE, DUFLO, 2009, 2011). Si les expérimentations aléatoires de terrain (ou *Randomized Controlled Trials* [RCT]) ne constituent qu'une part faible (en nombre) bien qu'importante (en budget) des expérimentations ainsi conduites dans le cadre du FEJ⁽¹⁾, elles méritent une attention toute particulière par leur nouveauté en France et par les questions spécifiques qu'elles posent (contraintes liées au protocole, rapports entre le scientifique et le politique, modalités de généralisation des résultats, etc.). Le développement de ces expérimentations a été particulièrement important dans les domaines du travail et de l'emploi, notamment sur le champ de l'insertion. Il est bien sûr encore trop tôt pour détecter les éventuelles évolutions que ce fonds a pu favoriser, mais il s'agit là d'un observatoire de choix. Les centaines d'expérimentations qui ont été financées à cette occasion préfigurent pour certains (FERRACCI, WASMER, 2012) les voies que les politiques seront susceptibles d'emprunter dans l'avenir.

Sur le plan international, les politiques du travail et de l'emploi ont été parmi les plus étudiées et influencées par ce type de méthodes, comparativement à l'éducation par exemple où la tradition des

expérimentations randomisées est plus restreinte⁽²⁾. Au-delà de l'actualité récente, les expérimentations par assignation aléatoire s'inscrivent dans une histoire longue et pluridisciplinaire (médecine, psychologie, sociologie, économie, etc.)⁽³⁾. Appliquées aux politiques publiques, elles avaient suscité un enthousiasme nourri aux États-Unis dans les années 1960-1970. Le projet d'une « société de l'expérimentation » (MONNIER, 1992) s'était traduit par la mise en œuvre de dispositifs mobilisant des milliers de personnes et des sommes considérables (des centaines de millions de dollars), notamment en matière de politique de l'emploi (ALLÈGRE, 2008). Dans les années 1980, l'essor initial a laissé place à un relatif reflux pour retrouver aujourd'hui un nouvel élan parmi les économistes.

En se focalisant plus particulièrement sur le contexte français récent, ce numéro a un double objectif : mettre en évidence des contributions issues d'économistes ayant mis en œuvre des expérimentations par assignation aléatoire sur le champ du travail et de l'emploi (les deux premiers articles) et prendre ces expérimentations pour objet d'étude en posant un regard réflexif, sociologique et juridique, sur ces approches (les deux derniers articles). Nous espérons ainsi faire dialoguer ces différentes disciplines autour d'une méthodologie qui traverse les sciences sociales. En introduction, nous présentons à grands traits les apports et limites actuelles des expérimentations aléatoires de terrain avant d'interroger les dimensions scientifique et politique de ces méthodes et de détailler le contenu des contributions constitutives de ce numéro.

(*) Les auteurs remercient vivement Thomas Amossé, Thierry Kirat et Magali Marcille pour leurs suggestions et leur relecture attentive.

(**) Centre de recherche sur l'industrie, les institutions et les systèmes économiques d'Amiens (Criisea), Université de Picardie-Jules-Verne ; agnes.labrousse@u-picardie.fr

(***) Direction de l'animation de la recherche, des études et des statistiques (Dares), Centre de recherche en économie et en statistique (Crest) ; philippe.zamora@dares.travail.gouv.fr

(1) La plus grande part des expérimentations conduites était menée sur un petit nombre d'individus et évaluée par des protocoles d'observation qualitative. Elles se distinguent donc des RCT qui cherchent à quantifier l'impact causal sur les publics bénéficiaires.

(2) Voir par exemple Cook, 2002.

(3) Ces thématiques ont été explorées lors des journées d'études « Le retour de la société de l'expérimentation ? Perspectives historiques et interdisciplinaires », qui se sont tenues les 25 et 26 novembre 2011 à Lyon, <http://calenda.org/206177> ; consulté le 15 septembre 2013.

Apports et spécificités des expérimentations de terrain

Les expérimentations aléatoires de terrain sont souvent désignées sous le nom d'expérimentations contrôlées ou randomisées : il s'agit d'une technique issue des essais cliniques utilisés dans le domaine médical et visant à sélectionner des groupes par tirage au sort. L'un de ces groupes est exposé à la politique évaluée (groupe traité), l'autre n'y est pas exposé ou, tout au moins, dans une moindre mesure que le groupe traité (groupe témoin). L'impact est calculé à partir de la comparaison du devenir des deux groupes.

Cet instrument d'évaluation présente une double ambition : d'une part, celle de mettre en évidence et quantifier l'effet causal moyen d'une intervention donnée (programme social, politique publique) sur un comportement ou un état (retour à l'emploi, amélioration de la santé, *etc.*); d'autre part, celle d'assurer, de viser, une forte validité externe des résultats obtenus (puisque ces expérimentations se déroulent dans un cadre réel et dans des conditions aussi proches que possible du cadre habituel de vie des acteurs). À ce titre, elles se distinguent des expérimentations conduites en laboratoire qui ont pour objectif d'étudier *in vitro* les mécanismes de décision des agents, à travers des jeux modélisant une situation économique. En comparant les comportements de deux groupes statistiquement similaires – et à condition que les populations expérimentées soient suffisamment nombreuses compte tenu de la précision souhaitée et de l'ampleur de l'effet –, elles mettent en évidence très simplement la *plus value nette* d'une intervention ou d'un programme. Ces méthodes permettent *a priori* de corriger la «sélection endogène», inhérente à la plupart des politiques publiques. Par ce phénomène aujourd'hui bien connu, les bénéficiaires d'un programme se distinguent des non-bénéficiaires par un grand nombre de caractéristiques bien souvent inobservables. La simplicité technique (écart des moyennes des deux groupes et calculs de précision) n'est pas le moindre des apports de ce type de méthodes. Le récent regain académique des expérimentations de terrain est de ce point de vue une réaction également à la technicité croissante de l'économétrie, qui avait entraîné les économistes vers des débats méthodologiques internes et avait mis au second plan les résultats proprement dits. Parmi les RCT, le *testing* à l'embauche est un outil bien connu, et utilisé même au-delà du milieu académique : il s'est considérablement développé en France ces dernières années, comme le rappelle Clémence Berson dans ce numéro.

Les expérimentations aléatoires de terrain sont-elles pour autant la panacée tant attendue dans le champ de l'évaluation des politiques publiques ? En

réalité, si la sélection aléatoire des groupes résout la difficulté fondamentale du «contrefactuel», surgissent des questions complexes liées à la nature *expérimentale* des programmes évalués par ce type de méthode. Il est rare en effet (mais pas impossible⁽⁴⁾) que ce type de méthodes s'applique à des programmes de droit commun, installés par la loi et accessibles à tous.

Les questions de validité externe des expérimentations

Dans quelle mesure un programme expérimental peut-il apporter des enseignements universels ou généralisables ? Il s'agit là d'une question fondamentale que le FEJ (dont la genèse et le fonctionnement sont retracés dans ce numéro par Marie-Christine Bureau et ses coauteurs), dans ses multiples appels à projets, a posée explicitement. Il est difficile d'y répondre dans l'absolu, tant les conditions concrètes des expérimentations peuvent être diverses et faire varier par là même la solidité de leur validité externe. Comme le rappellent énergiquement Bernard Gomel et Évelyne Serverin (également dans ce numéro), il s'agit là d'une des critiques les plus fréquentes adressées aux expérimentations (aléatoires ou non), mais il nous semble pourtant que cette critique ne peut s'appliquer uniformément à l'ensemble de ces travaux.

Contrairement aux essais cliniques dans le domaine médical, le principe du double aveugle est impossible en sciences sociales. Les groupes «traité» et «témoin» sont nécessairement dans un rapport différent au caractère expérimental du programme évalué. Le fait que le programme soit expérimental, et donc exceptionnel, peut être connu – à des degrés divers – par les individus du groupe traité comme par ceux du groupe témoin. Aux évaluateurs de l'expérimentation revient la tâche de montrer que ce caractère «expérimental» n'a pas modifié artificiellement le comportement des uns et des autres et surtout affecté les deux groupes dans un sens différent. C'est là une des étapes importantes d'un travail d'évaluation sur données expérimentales, mais non nécessairement une limite indépassable, contrairement, il nous semble, à ce qu'affirment Gomel et Serverin. Ainsi, dans les expérimentations importantes sur les opérateurs privés de placement (OPP) conduites par l'ANPE, l'Unedic et le ministère du Travail (BEHAGHEL *et al.*, 2012), les programmes évalués n'étaient pas d'une nature exceptionnelle en soi, et les demandeurs d'emploi n'étaient pas vraiment conscients de l'évaluation en cours. La validité externe des résultats obtenus – consolidés

(4) On peut citer en particulier les récentes évaluations américaines des programmes de formation pour les jeunes non qualifiés : Job Corps (voir SCHOCHET *et al.*, 2008).

par leur constance et leur robustesse – avait toute chance d’être forte au moins de ce point de vue.

D’autres incertitudes pèsent bien sûr sur la possibilité de généraliser les résultats d’une évaluation expérimentale à d’autres publics ou à d’autres lieux. Cette critique est souvent citée pour les études expérimentales. Pourtant, elle s’applique à n’importe quelle étude empirique : les résultats ne doivent être évidemment considérés qu’en lien avec les conditions concrètes de son déroulement. Aux méta-études le soin de se prononcer sur une possible généralisation. Les « randomisateurs » le savent bien, qui se livrent, lorsqu’ils le peuvent, à des répliques de leurs évaluations sur d’autres terrains et d’autres publics. Néanmoins, dans ce même ordre d’idées, Gomel et Serverin (*op. cit.*) rappellent un aspect qui pose plus spécifiquement question dans le cas des expérimentations. Dans le cas de *designs* expérimentaux simples, généralement testés sur un petit nombre de personnes, les expérimentations sont souvent impuissantes à détecter les effets externes des programmes. En d’autres termes, le programme est susceptible d’avoir des effets non seulement sur les individus bénéficiaires mais aussi sur les non-bénéficiaires. Un tel effet externe s’amplifie lorsque le nombre d’individus traités s’accroît. En cas de généralisation, l’impact réel a toutes les chances de différer fortement de l’impact mesuré lors de l’expérimentation, il peut être fortement atténué ou amplifié, selon le sens des effets externes. Même s’il s’agit là d’une limite qui est commune aux différents types d’expérimentation, le cadre rigoureux des RCT et l’estimation quantifiée qu’elles proposent sont susceptibles de favoriser l’occultation des difficultés associées à une éventuelle généralisation. Un très bel article décrit un cas tout à fait emblématique de ce phénomène : Amy FINKELSTEIN (2007) montre ainsi que les effets du programme d’assurance-maladie *Medicare* adopté aux États-Unis en 1967 pour les personnes âgées sont en réalité bien supérieurs à ce que suggérait l’expérimentation aléatoire (*Rand Health Insurance Experiment*) conduite de 1974 à 1982 dans ce même pays. Exploitant les données de l’activité de l’ensemble des hôpitaux américains, dans une approche relevant plutôt de l’expérience naturelle, FINKELSTEIN montre que l’adoption du *Medicare* a contribué à transformer profondément l’offre et la technologie de soin et a donc eu des impacts généralisés sur l’ensemble des individus, qu’ils soient assurés ou non.

Enquêtes qualitatives et quantitatives : des ponts à construire

Les expérimentations par assignation aléatoire déploient une approche principalement quantitative. Néanmoins, se développent aujourd’hui des combinaisons entre expérimentations de terrain et études de terrain qualitatives (observation directe,

entretiens semi-directifs), à l’instar de l’expérimentation présentée par Marc Ferracci et Florine Martin dans ce dossier.

En quoi les deux dimensions qualitatives et quantitatives peuvent-elles être sources de complémentarités ? On peut avancer ici quelques éléments. Une étude qualitative exploratoire peut contribuer à définir le *design* expérimental (choix des traitements pertinents, des populations cibles, adéquation des unités et des catégories statistiques et conceptuelles au terrain⁽⁵⁾), en faisant remonter des informations sur le contexte local, sur les contraintes et variables pertinentes pour les acteurs, en anticipant les facteurs de résistance. Dans la phase de mise en œuvre puis d’interprétation de l’expérimentation, l’enquête qualitative peut permettre d’identifier des écarts dans l’observation du protocole *in situ*, de mieux cerner des effets non intentionnels, des facteurs de blocage mais surtout de mieux repérer les chaînes causales par l’observation des processus et en interrogeant les acteurs sur leurs actions, leurs motifs et leurs contraintes pratiques. Car, si les expérimentations randomisées permettent de mesurer rigoureusement l’effet moyen d’un programme, les processus causaux à l’œuvre (pourquoi, par quels mécanismes, ces effets ont pris place ?) peuvent être délicats à mettre en évidence. Comme la plupart des évaluations quantitatives, les expérimentations ne permettent pas d’identifier directement le « chemin » de la causalité. D’où l’intérêt de croiser les résultats des RCT avec d’autres matériaux (JATTEAU, 2013), comme les monographies sur l’emploi des femmes dans la grande distribution que mobilise Clémence Berson, dans ce numéro, pour mieux saisir les motifs de la discrimination à l’embauche qu’elle identifie dans l’expérimentation qu’elle a conduite.

Néanmoins, la combinaison des deux méthodologies n’est pas aisée. Si les études qualitatives sont souvent peu coûteuses, les obstacles sont souvent autres que financiers. Bien que toutes deux soumises à des règles de scientificité, ces deux modalités d’administration de la preuve relèvent de cultures scientifiques (et souvent d’équipes de recherche) distinctes. Cela peut donner lieu à des incompréhensions, des tensions et des rendez-vous manqués, comme l’ont mis en évidence Aurélie QUENTIN et Isabelle GUÉRIN (2013) dans le cas de l’évaluation du programme de microassurance Sky au Cambodge.

(5) On peut penser aux travaux du Groupe Amira (Amélioration des méthodes d’investigation en milieu rural africain). Créé en 1975, ce groupe de travail pluridisciplinaire s’est penché sur l’utilisation de méthodes à la fois quantitatives et qualitatives, le « qualitatif » donnant sens au « quantitatif » au moment de l’usage, de la phase interprétative du chiffre mais aussi, en amont, dans la conception même du *design* statistique (COUTY, 1983).

Les expérimentations à la croisée des logiques scientifiques et politiques

Les expérimentations randomisées contrôlées sont, pour reprendre une belle formule d'Alain DESROSIÈRES (2008), à la fois des outils de preuve et de gouvernement. Ces deux dimensions, nous le verrons, peuvent entrer en tension et nécessiter de multiples opérations de traduction pour passer du registre scientifique au registre politique. La possibilité même de conduire des expérimentations en France a nécessité une évolution du paysage juridico-administratif français.

Un instrument d'évaluation plongé dans le contexte juridico-politique français

L'appel à contributions qui est à l'origine de ce numéro encourageait, entre autres choses, des comparaisons internationales. Les articles qui le composent se focalisent finalement sur le contexte français mais fournissent, ce faisant, de précieuses indications sur l'évolution du contexte de l'évaluation des politiques publiques en France. D'autres formes d'évaluation y ont longtemps occupé le devant de la scène, où les premiers experts de l'évaluation des politiques publiques étaient largement rétifs au modèle de l'expérimentation aléatoire (MONNIER, 1992).

L'introduction en France de ces méthodes est donc relativement tardive comparativement aux pays anglo-saxons. Leur importation est passée par des « entrepreneurs institutionnels⁽⁶⁾ » issus des mondes académiques et politico-administratifs ayant une excellente connaissance des pratiques nord-américaines. On peut ainsi mentionner deux économistes français d'envergure internationale, en prise directe avec les États-Unis : Esther Duflo, professeure au Massachusetts institute of technology (MIT) et qui entretient des liens forts avec l'École d'économie de Paris ; François Bourguignon, qui a promu de telles évaluations à la Banque mondiale et a présidé l'École d'économie de Paris ainsi que le comité d'évaluation du revenu de solidarité active (RSA). Sur le plan institutionnel, Martin Hirsch, lui aussi familier des

expérimentations nord-américaines⁽⁷⁾, a assurément joué un rôle décisif, de la commission « Familles, vulnérabilité, pauvreté » de 2005 à sa participation à l'ouverture gouvernementale sous la présidence Sarkozy. Son arrivée au Haut Commissariat aux solidarités actives contre la pauvreté puis au Haut Commissariat à la jeunesse et son départ vont largement conditionner la création et le devenir des expérimentations, comme le mettent en évidence Bureau et ses coauteurs dans ce numéro.

L'introduction des expérimentations aléatoires a nécessité des transformations du cadre juridique et administratif français. Il a progressivement été aménagé pour permettre ce type d'évaluation et les intégrer à l'arsenal des politiques publiques, une évolution retracée dans les deux derniers articles de ce numéro. Comme le montrent fort bien Bureau et ses coauteurs, les premières expérimentations s'inscrivent dans une reconfiguration des rapports entre l'État et les collectivités locales.

Projet politique et projet scientifique : des discordances potentielles dans les conceptions, les temporalités et les modes de généralisation des expérimentations

Les expérimentations impliquent des acteurs diversifiés : administrations publiques, élus, organismes de recherche, associations, entreprises privées et cabinets de conseil, « sujets » des expérimentations. Les relations qu'ils entretiennent sont redessinées lors des programmes expérimentaux et leurs rapports aux expérimentations sociales ne sont pas nécessairement concordants. Deux groupes d'acteurs clés retiennent tout particulièrement l'attention : les acteurs politico-administratifs à l'échelle nationale ou locale ; et les experts scientifiques, généralement issus du milieu académique. Comme le soulignent Gomel et Serverin dans ce numéro, « les “randomisateurs” apportent la promesse d'une plus grande certitude scientifique quant aux effets d'un dispositif sur des comportements donnés » (p. 61). Les expérimentations

(7) Sensibilisé aux essais cliniques par sa formation médicale, il en connaissait les applications en sciences sociales aux États-Unis par l'entremise de sa cousine Judith Gueron, qui fut directrice de l'institut de recherche *Manpower development research corporation* (MDRC), un organisme « fer de lance » dans la mise en œuvre des RCT aux États-Unis (voir Chelle E. [2011], « “Comparing apples and oranges?” La question de la comparabilité de la France et des États-Unis au prisme de deux dispositifs sociaux », in *Comparer en sciences sociales : une science inexacte?*, École thématique internationale CNRS-PACTE/LATTS/EPFL Grenoble, jeudi 1^{er} et vendredi 2 décembre 2011. Disponible en ligne à l'adresse : <http://hal.archives-ouvertes.fr/docs/00/64/05/93/PDF/TexteET2011Chelle.pdf>; consulté le 20 août 2013). Judith Gueron est intervenue aux côtés de Martin Hirsch lors du grand colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation des 22 et 23 mai 2008 (voir la retranscription en ligne à l'adresse : http://travail-emploi.gouv.fr/IMG/pdf/Retranscription_ADM_V2.pdf; consulté le 20 août 2013).

(6) Devaux-Spatarakis A. (2012), « The naturalization of RCTs in France. Story of “the experimental method” », article présenté au colloque international AHE-IIPPE-FAPE *Political economy and the outlook for capitalism*, session « Randomized controlled experiments, economics and social sciences », Université Paris 1 Panthéon-Sorbonne, 5-7 juillet 2012.

sociales permettraient de trancher par la force de la preuve scientifique, par une analyse scientifique rationnelle, dépassionnée, des débats politiques à forte composante idéologique (COMMISSION FAMILLE, VULNÉRABILITÉ, PAUVRETÉ, 2005 ; BANERJEE, DUFLO, 2011), promesse qui était aussi celle de l'*experimenting society* à la fin des années 1960 (MONNIER, 1992). Les articles de ce dossier offrent une vision plus complexe et plus riche des interactions entre le politique et le scientifique qui s'engagent dans l'expérimentation sociale, des contraintes comme des ressources qui s'y jouent pour les différents groupes d'acteurs.

Le FEJ a représenté une manne financière importante aussi bien pour les acteurs politiques que pour les centres de recherche pratiquant les expérimentations. Néanmoins, au-delà de cet intérêt commun, la définition de ce qui fait un « bon projet » à expérimenter peut diverger selon que l'on se place du point de vue des acteurs politiques ou des chercheurs.

Pour les acteurs locaux, un bon projet doit répondre aux spécificités locales comme aux attentes de l'ensemble des acteurs engagés et peut donc combiner un nombre important de mesures. Ce type de projets relève – selon les termes d'un ancien membre du FEJ rapportés par Bureau et ses coauteurs – d'un projet « sapin de Noël » (p. 51). Ils ne satisfont pas aux critères scientifiques d'une expérimentation randomisée contrôlée : foisonnants, ils ne sont guère adaptés à une expérimentation aléatoire car ils nécessiteraient des effectifs et des moyens pléthoriques avec une démultiplication des bras de l'expérimentation. Comme les essais cliniques (LABROUSSE, 2010), les expérimentations contrôlées se prêtent mal au test des effets de traitements combinés (au-delà de deux ou trois molécules). Pour isoler rigoureusement l'effet propre d'une mesure, les expérimentations randomisées contrôlées impliquent généralement des dispositifs plus rudimentaires du point de vue des décideurs locaux. Jean-David NAUDET et ses coauteurs (2012), qui ont travaillé ces questions à l'Agence française de développement (AFD), qualifient ce genre de projets ajustés à la randomisation de « projet-tunnel »⁽⁸⁾ : un projet où le lien causal entre le traitement et son effet s'exerce de manière relativement rapide, simple, linéaire, sans effets récursifs, à l'instar des vermifuges visant à réduire l'absentéisme en classe

au Kenya (KREMER, MIGUEL, 2004⁽⁹⁾). En outre, plus le projet est ancré dans les caractéristiques locales, plus la question épineuse de la validité externe évoquée plus haut se pose et les possibilités de généralisation du dispositif testé s'amenuisent. Or il s'agit là d'une préoccupation importante pour les acteurs académiques et pour l'État central soucieux de développer des projets à l'échelle nationale. Pour le dire de manière imagée, le « projet-tunnel » brille d'autres lumières que le « projet-sapin de Noël ».

Du point de vue académique, une bonne expérimentation sociale, c'est celle qui respecte de bout en bout la rigueur statistique du protocole, contrôle les biais potentiels de l'expérimentation (biais de randomisation, d'attrition, de substitution, *etc.*) et se prolonge sur une durée suffisante. Ces exigences peuvent apparaître aux acteurs politiques, privés et associatifs locaux, dont la culture statistique est souvent réduite, comme des rigidités paralysantes. Le protocole des expérimentations randomisées nécessite une organisation relativement centralisée, laissant dans son déroulement peu de marges aux parties prenantes locales, même si quelques ajustements peuvent prendre place en réponse à des difficultés opérationnelles comme le montre l'article de Ferracci et Martin dans ce numéro. Les acteurs locaux peuvent se sentir dépossédés des mesures testées ; il faut alors que les experts scientifiques gèrent la frustration voire le désintérêt des premiers envers les expérimentations en cours. À cet égard, un parallèle avec les essais cliniques peut être esquissé. Nicolas DODIER et Janine BARBOT (2000) ont mis en lumière, à propos des essais thérapeutiques sur le Sida, des « tensions épistémiques » entre, d'un côté les promoteurs des grands essais antiviraux répondant aux canons des RCT, essais centralisés avec une organisation hiérarchisée et planifiée, d'une durée longue, portant sur un grand nombre de patients et, de l'autre, les partisans d'essais « souples » (associations de patients, certains cliniciens) reposant sur une organisation distribuée, avec des *designs* plus en prise avec les questions émergentes que se posent les cliniciens mais mettant en péril la robustesse de l'essai. Ce qui peut apparaître comme un effet d'apprentissage local pour les praticiens constitue une perturbation de l'expérimentation pour les autres. Dans certains cas, les acteurs locaux parviennent à infléchir le devenir de l'expérimentation et à en faire évoluer les objectifs, comme le notent Gomel et Serverin à propos d'une

(8) NAUDET et ses coauteurs (2012, p. 13) détaillent ainsi les prérequis des RCT : «(i) une période cohérente avec la chaîne causale supposée, (ii) un nombre limité de traitements, homogènes et précis, (iii) un processus d'administration préalablement testé, (iv) une chaîne causale courte et indépendante d'événements extérieurs, (v) une appropriation rapide et stable dans le temps par les bénéficiaires du traitement, (vi) une participation large et stable dans le temps des bénéficiaires, et (vii) un ensemble d'effets mesurables à court ou moyen terme, qui couvre les principaux aspects du traitement» (traduction de la revue).

(9) Les vers intestinaux – touchant un enfant sur quatre dans le monde – provoquent fatigue chronique et anémie et peuvent être soignés par la simple prise d'un comprimé vermifuge tous les six mois. L'expérimentation, reposant sur la sélection aléatoire d'écoles entières, montre l'amélioration rapide de l'état de santé des enfants traités, amélioration qui se traduit par une nette diminution de l'absentéisme en classe. Il s'agit donc d'un dispositif élémentaire, avec des effets rapides, où la chaîne causale est courte, simple et univoque (le vermifuge améliore l'état de santé des enfants et, par conséquent, leur présence) et bien identifiée par la littérature médicale.

expérimentation sur le placement des chômeurs. Au-delà de la méthode, la discordance peut porter sur le choix même des programmes faisant l'objet d'expérimentations. Ce fut le cas d'un dispositif de réduction de l'absentéisme scolaire dans l'académie de Créteil (la « cagnotte scolaire »), où le programme défini par des chercheurs de l'École d'économie de Paris a rencontré des résistances telles que le projet a été interrompu après la phase pilote (Gomel et Serverin ; Bureau *et al.* dans ce numéro). On avait là une expérimentation tout à fait légitime sur le plan des théories économiques standard – modifier le système d'incitations (collectif) des élèves en introduisant des récompenses financières – mais elle est apparue comme illégitime à de nombreux acteurs sociaux et politiques français.

On retrouve de tels déphasages dans les calendriers et les agendas thématiques des différents acteurs. Ainsi, comme le montrent Bureau et ses coauteurs, le calendrier politique des deux premières années d'existence du FEJ, conjugué au temps compté du haut commissaire d'ouverture Martin Hirsch, a imprimé de son urgence le travail des chercheurs : rythme de travail extrême, succession des appels à projets, urgence qui n'a été supportable qu'à la faveur d'un « esprit du livre vert » et grâce à l'enthousiasme des pionniers. Car cette impatience politique est peu en accord avec un calendrier scientifique plus lent, qui nécessite de disposer du temps nécessaire pour apprendre des premiers essais engagés et les laisser aboutir. Réciproquement, cette temporalité scientifique peut être en décalage avec le rythme politique : certaines expérimentations s'achèvent au moment où les acteurs politiques les ayant initiées ne sont plus chargés des dossiers (remaniements, alternance politique), ce qui met en péril l'appropriation politique des résultats. Ces observations ne sont pas spécifiques au cas français : elles avaient déjà été faites dans le cas nord-américain dès les années 1980 (MONNIER, 1992, pp. 44-45).

Quant à la possible généralisation des résultats, la notion même revêt des sens différents selon que l'on se situe dans le registre politique ou scientifique. Comme le relèvent GOMEL et SERVERIN (2009, p. 14), « d'un côté, la généralisation est une opération intellectuelle d'un chercheur qui est ou non autorisé à étendre ses conclusions au-delà du champ étroit de son expérimentation ; de l'autre, la généralisation est l'opération institutionnelle par laquelle une disposition prise localement [...] fait l'objet d'une loi applicable à l'ensemble du territoire ». La première renvoie à la question scientifique de la validité externe quand la seconde correspond à un mode de validation politique et un changement d'échelle administrative. En pratique, les contraintes différenciées des acteurs expliquent que « la généralisation des expérimentations, point nodal de la relation entre innovation locale, expertise et décision publique, est restée, à des titres divers, un point

aveugle pour l'ensemble des acteurs embarqués dans l'expérience » (Bureau *et al.* dans ce numéro, p. 41).

Les résultats des expérimentations ne suffisent pas en eux-mêmes à déterminer des politiques publiques idoines. De multiples processus de « traduction » sont nécessaires pour passer des résultats statistiques à un programme politique (et ce, même en faisant abstraction des processus parlementaires et administratifs). En effet, l'interprétation des résultats est souvent délicate, et parfois incertaine. Les précautions scientifiques dont s'entourent légitimement les chercheurs peuvent être une source d'embarras pour des politiques à la recherche de préconisations univoques. Symétriquement, les « randomisateurs » sont attendus dans un rôle auquel, particulièrement lorsqu'ils font partie du monde académique, leur pratique ne les prépare en général pas : dire ce qui « marche » ou ne « marche pas », voire s'engager dans des préconisations de généralisation, de transformation ou de suppression des dispositifs expérimentaux⁽¹⁰⁾. Les arbitrages politiques entre mesures ne peuvent se faire sur la seule base des expérimentations : « les résultats des évaluations ne suffisent pas pour arbitrer entre plusieurs mesures évaluées positivement » (Bureau *et al.*, p. 52). En particulier, un programme peut être efficace mais coûteux, ce qui soulève la question du passage de la mesure des effets à une analyse coût-bénéfice. Ce point est clairement explicité dans le rapport d'activité du FEJ mentionné par Bureau et ses coauteurs (p. 53) et, dans leur contribution, Ferraci et Martin, en font l'une des deux directions à explorer dans un futur proche « en produisant une véritable analyse coût-bénéfice de chacun des dispositifs » (p. 26).

Quelles frontières et quelles circulations entre les registres politiques et scientifiques ?

Dans plusieurs cas, les expérimentations ont été instrumentalisées par les acteurs politiques, en décalage avec les normes de scientificité des chercheurs. Dans le cas du RSA, l'expérimentation n'a pas eu lieu selon les canons de l'expérimentation aléatoire. Comme le relèvent Gomel et Serverin dans ce numéro, cela n'a pas empêché un usage politique sélectif de résultats pourtant discutables et de moins en moins significatifs au cours du temps. Peuvent alors apparaître des tensions « lorsque le décideur politique adopte une nouvelle mesure en préemptant les résultats de l'expérimentation » (Bureau *et al.*, p. 52). Dans des cas extrêmes, les

(10) Ici aussi, les observations de la situation française récente font largement écho à l'expérience nord-américaine des années 1970-1980 : « Les commanditaires [politiques], déçus par l'incapacité des universitaires à formuler des conclusions en terme d'implications politiques plus larges, ont renoncé à la mise en place de dispositifs lourds étroitement focalisés sur un seul objectif » (MONNIER, 1992, p. 45).

impératifs politiques peuvent être en porte-à-faux complet avec les exigences de scientificité des expérimentations. Ainsi des projets d'expérimentations dans les territoires d'outre-mer dont la qualité scientifique apparaissait discutable ont-ils été validés pour des raisons politiques (crise en 2009 en Guadeloupe, taux de suicide des jeunes en Guyane, *etc.*), en passant outre l'avis du conseil scientifique du FEJ, comme le met en évidence la contribution de Bureau et de ses coauteurs. Ces expérimentations « politiques » constituent un sujet de tensions récurrentes entre le scientifique et le politique.

Quelle attitude adopter alors face au « brouill[age] [de] la frontière entre recherche et décision politique » (Gomel et Serverin, p. 68), face à « l'enchevêtrement inextricable des registres politique et scientifique » (Bureau *et al.*, p. 53)? Il est en effet choquant que l'autonomie du « savant » dans ce qui est au cœur de sa compétence – la définition d'un protocole robuste – ne soit pas respectée par les politiques. Pour Gomel et Serverin, il est alors souhaitable que l'expérimentation aléatoire retrouve sa place dans les laboratoires, loin des promesses de l'ingénierie sociale. Cette étanchéité entre expérimentations scientifiques et administratives qu'ils appellent de leurs vœux soulève peut-être autant de questions qu'elle apporte de réponses. N'est-ce pas là reconduire une rhétorique de neutralité idéologique ou axiologique de la science caractéristique de l'ingénierie sociale, rhétorique remise en question par les auteurs de l'article? Ces derniers assignent le contrôle de la pertinence des expérimentations « aux communautés scientifiques ». Les décideurs politiques – avec leur légitimité démocratique, aussi fragile et incomplète soit-elle – n'ont-ils donc rien à dire sur la pertinence des « traitements » évalués? Des expérimentations commandées par des préoccupations scientifiques, c'est-à-dire des préoccupations théoriques inscrites dans des moments et des paradigmes particuliers des sciences sociales, sont-elles nécessairement gages de pertinence pratique pour les « sujets traités », à savoir les citoyens (ou certaines catégories parmi eux)? En outre, la division du travail entre les différentes sciences sociales prônée par les auteurs peut à son tour interroger. Est-il toujours simple de définir clairement le champ disciplinaire sachant que le social est nécessairement pluridimensionnel et que la division académique du travail, elle-même socialement et historiquement construite, peut être un obstacle à la compréhension des phénomènes sociaux dans leur complexité?

Ne pourrait-on pas, au contraire, plaider pour des forums « tripartites » ou « hybrides » (comptant des politiques, des scientifiques et des citoyens) dont les modalités de fonctionnement assureraient un dialogue plus équilibré entre préoccupations politiques, scientifiques et civiques, en incluant également les futurs sujets des expérimentations, et ce, aussi bien dans la définition des traitements

que dans l'interprétation des résultats (pour limiter les instrumentalisations politiques comme scientifiques)? La question de la place des « sujets » de l'expérimentation va de pair avec celle de l'éthique, tout particulièrement en présence de « sujets vulnérables », un enjeu fondamental dont Gomel et Serverin évoquent les principaux tenants et aboutissants dans leur conclusion. Nous n'entendons pas apporter ici de réponses à ces questions éminemment complexes, mais un des mérites des articles qui composent ce numéro nous semble être d'avoir contribué à les formuler.

Parcours de lecture

L'article de Marc Ferracci et Florine Martin est emblématique des recherches conduites sur les politiques actives du marché du travail, qui cherchent à identifier les pratiques les plus efficaces ou les plus efficaces en termes d'impact sur le retour à l'emploi. Il s'inscrit dans la lignée des travaux récents sur l'accompagnement, qui ne sont pas forcément expérimentaux (voir, entre autres, les travaux de l'IAB – l'Institut de recherche sur l'emploi allemand – et ceux de l'équipe de Michael Rosholm de l'université d'Aarhus). Les auteurs ne s'intéressent ni à la motivation des demandeurs d'emploi, ni à leur contrôle, mais aux méthodes de travail des conseillers pour l'emploi. Ils étudient précisément deux types de changements dans l'organisation du travail d'un opérateur de placement (la Sodie) : le premier vise à améliorer la gestion du temps de travail des conseillers ; le second met davantage de moyens au service de la prospection des emplois. D'après le programme expérimental suivi, les deux dispositifs accroissent les chances de retour à l'emploi. Les auteurs soulignent ce résultat mais ne se prononcent pas sur les éventuels effets externes des dispositifs, dont la détermination conditionne fortement l'intérêt de leur généralisation.

Le deuxième article, proposé par Clémence Berson, s'appuie sur une forme particulière d'expérimentation aléatoire – le *testing* – qui s'est considérablement développée en France au cours des dernières années et a contribué à mettre en évidence l'existence de discriminations à l'embauche à l'égard des travailleurs issus de l'immigration. L'auteure trouve des résultats similaires dans le secteur de la grande distribution mais sa contribution essentielle est de chercher à comprendre si les discriminations mises en évidence sont économiquement rationnelles (discrimination « statistique ») ou non (discrimination « par goût »). Elle propose de trancher entre les deux types de discriminations en observant l'évolution de la discrimination en fonction du degré de concurrence existant localement entre les super- et hypermarchés. Suivant l'hypothèse que la discrimination statistique a

tendance à diminuer si la concurrence augmente alors que la discrimination par goût a tendance à rester constante, elle conclut que la discrimination envers les personnes ayant un patronyme marocain se fait «par goût». Même si l'article ne permet pas totalement de trancher, les différences n'étant pas toujours significatives, il suit une piste originale pour interpréter les discriminations mises en évidence dans un cadre expérimental.

Suivent deux articles de mise en perspective de l'essor récent des expérimentations de terrain en France. Dans le premier, dans une perspective de sociologie pragmatique, au plus près des acteurs, Marie-Christine Bureau, François Sarfati, Jules Simha et Carole Tuchsirer portent leur regard sur les «acteurs embarqués» dans les expérimentations de terrain. Ils montrent comment l'histoire des expérimentations sociales à la française prend sens à la lumière du jeu triangulaire entre administrations centrales, acteurs locaux et experts. Après avoir retracé l'évolution ambiguë du contexte juridique et politique français des années 2000, les auteurs analysent notamment la succession des appels à projets pour montrer qu'après une phase initiale visant à faire remonter des solutions locales innovantes sur une gamme étendue de problèmes sociaux, le FEJ s'est recentré sur les thématiques

éducatives et a été l'objet d'une reprise en main par l'État. Pour les auteurs, la question de la généralisation s'avère un point aveugle pour les différents groupes d'acteurs.

Plus normatif et résolument critique, l'article de Bernard Gomel et Évelyne Serverin, questionne l'expérimentation sociale aléatoire selon trois lignes argumentatives : l'inscription juridique des expérimentations, les canons scientifiques mis à l'épreuve des pratiques expérimentales et l'apport des expérimentations aux décisions politiques. L'article resitue tout d'abord les évolutions juridiques dans lesquelles s'enchaînent les expérimentations pour ensuite montrer, à partir d'exemples choisis, en quoi les situations pratiques peuvent s'écarter des attendus de l'expérimentation scientifique (maîtrise de la conduite de l'expérimentation, isolation de l'effet propre). Ils discutent finalement de la contribution des expérimentations aux réformes, en insistant sur les difficultés à transposer dans le registre politique les résultats des expérimentations dont l'objectif s'infléchit peu à peu dans les textes : visant initialement à assurer une «portée démonstrative forte» aux réformes, ils contribuent aujourd'hui de manière plus classique à apporter des «éléments de réflexion» (p. 67) à celles-ci.

Bibliographie

ALLÈGRE G. (2008), «L'expérimentation sociale des incitations financières à l'emploi : questions méthodologiques et leçons des expériences nord-américaines», *Document de travail de l'OFCE*, n° 2008-22.

BANERJEE A. V., DUFLO E. (2011), *Poor economics: a radical rethinking of the way to fight global poverty*, New York, Public Affairs.

BANERJEE A. V., DUFLO E. (2009), «L'approche expérimentale en économie du développement», *Revue d'économie politique*, vol. 119, n° 5, pp. 691-726.

BEHAGHEL L., CRÉPON B., GURGAND M. (2012), «Private and public provision of counseling to job-seekers: evidence from a large controlled experiment », *IZA discussion paper*, n° 6518.

COMMISSION FAMILLE, VULNÉRABILITÉ, PAUVRETÉ (2005), *Au possible nous sommes tenus. La nouvelle équation sociale : 15 résolutions pour combattre la pauvreté des enfants*, Paris, Ministère des Solidarités, de la Santé et de la Famille.

COOK T. D. (2002), «Randomized experiments in education policy research: a critical examination of the reasons the educational evaluation community has offered for not doing them», *Educational evaluation and policy analysis*, vol. 24, n° 3, pp. 175-199.

COUTY P. (1983), «Qualitatif et quantitatif», in Couty P., Winter G., *Qualitatif et quantitatif: deux modes d'investigation complémentaires. Réflexions à partir des recherches de l'ORSTOM en milieu rural africain*, pp. 35-47. Disponible en ligne à l'adresse : http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers4/15290.pdf; consulté le 28 septembre 2013.

DESROSIÈRES A. (2008), *L'argument statistique*, tome I et tome II, Paris, Presse de l'École des mines.

DODIER N., BARBOT J. (2000), «Le temps des tensions épistémiques. Le développement des essais thérapeutiques dans le cadre du Sida», *Revue française de sociologie*, vol. 41, n° 1, pp. 79-118.

FERRACCI M., WASMER E. (2012), *État moderne, État efficace : évaluer les dépenses publiques pour sauvegarder le modèle français*, Paris, Odile Jacob.

FINKELSTEIN A. (2007), «The aggregate effects of health insurance: evidence from the introduction of Medicare », *The quarterly journal of economics*, vol. 122, n° 1, pp. 1-37.

GOMEL B., SERVERIN E. (2009), «Expérimenter pour décider? Le RSA en débat», *Document de travail*, n° 119, Noisy-le-Grand, Centre d'études de l'emploi.

JATTEAU A. (2013), *Les expérimentations aléatoires en économie*, Paris, La Découverte, coll. «Repères».

KREMER M., MIGUEL E. (2004), « Worms: identifying impacts on education and health in the presence of treatment externalities », *Econometrica*, vol. 72, n° 1, pp. 159-217.

L'HORTY Y., PETIT P. (2010), «Évaluation aléatoire et expérimentations sociales», *Document de travail*, n° 135, Noisy-le-Grand, Centre d'études de l'emploi.

LABROUSSE A. (2010), «Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement», *Revue de la régulation* [en ligne], n° 7; <http://regulation.revues.org/index7818.html>; consulté le 28 septembre 2013.

MONNIER E. (1992), *Évaluations de l'action des pouvoirs publics. Du projet au bilan*, 2^e éd. rev. et augm., Paris, Economica.

NAUDET J.-D., DELARUE J., BERNARD T. (2012), « Impact evaluations : a tool for accountability ? Lessons from experience at AFD », *Proceedings of the 9th AFD-EUDN Conference*, pp. 25-43.

QUENTIN A., GUÉRIN I. (2013), «La randomisation à l'épreuve du terrain. L'exemple du projet de microassurance Sky au Cambodge», *Revue tiers monde*, n° 213, pp. 179-200.

SCHOCHET P. Z., BURGHARDT J., MCCONNELL S. (2008), « Does job corps work? Impact findings from the national Job Corps Study », *The American economic review*, vol. 98, n° 5, pp. 1864-1886.