

Guide méthodologique pour l'évaluation des expérimentations sociales

à l'intention des porteurs de projets

Ce guide méthodologique est destiné aux porteurs de projet qui souhaitent se lancer dans une expérimentation sociale. Son objectif est de présenter de façon claire les apports et les difficultés d'une démarche expérimentale lorsqu'elle est assortie d'un dispositif d'évaluation. La présentation des différentes approches de l'évaluation se veut aussi complète que possible, compte tenu de l'état de l'art à la fin des années 2000.

Deux grands types d'évaluations co-existent : les méthodes qui s'appliquent sur des échantillons représentatifs (de quelques centaines à quelques milliers d'individus, voire davantage), ou méthodes quantitatives, et les investigations approfondies sur des petits échantillons raisonnés de l'ordre de quelques dizaines d'individus, ou méthodes qualitatives. Au sein des méthodes quantitatives, les approches les plus fiables sont les méthodes d'expérimentation contrôlée et randomisée qui reposent sur l'observation d'un groupe test et d'un groupe témoin ne bénéficiant pas du dispositif, ces groupes étant construits en y affectant les personnes par tirage au sort. Lorsqu'une expérience randomisée ne peut être envisagée, d'autres méthodes quantitatives peuvent être mobilisées qui consistent à tenter de reproduire artificiellement les conditions d'une expérience randomisée à l'aide de techniques statistiques adaptées aux caractéristiques de l'expérimentation.

Au total, toutes les méthodes d'évaluation ne poursuivent pas les mêmes objets et ne répondent pas aux mêmes interrogations. L'évaluation mobilise un ensemble d'approches complémentaires qui permettent de répondre à des questions différentes. Ce guide est structuré autour de ces questions : Quels sont les objectifs visés par l'expérimentation ? Quels sont ses effets ? Comment les expliquer ?

INTRODUCTION

Définition d'une expérimentation sociale

C'est une innovation de politique sociale initiée dans un premier temps à petite échelle, compte tenu des incertitudes existantes sur ses effets et mise en oeuvre dans des conditions qui permettent d'en évaluer les résultats, dans l'optique d'une généralisation si ces résultats s'avèrent probants.

Les Rencontres de l'expérimentation sociale, Grenoble, novembre 2007

Lorsqu'une institution engage des ressources humaines et matérielles sur un programme social innovant, elle a besoin de savoir si ce programme améliore effectivement la situation des personnes qui en bénéficient. Cela suppose de pouvoir évaluer de façon rigoureuse les conditions de mise en oeuvre et les effets du programme sur ses bénéficiaires. Il lui est nécessaire en outre de savoir si un meilleur usage des ressources qu'elle mobilise peut être envisagé dans l'intérêt des personnes. Seule une évaluation rigoureuse permet de dire de façon fiable si le programme est efficace. Cette évaluation peut aussi renseigner sur les raisons pour lesquelles le programme fonctionne, pour réfléchir à la généralisation, ou au contraire fonctionne mal, pour pouvoir l'améliorer. L'objectif de l'évaluateur est de participer à l'accumulation de connaissances fiables et partagées sur l'efficacité des dispositifs en fonction des conditions de leur expérimentation. Expérimentation et évaluation ont donc tout à gagner à être associées de façon étroite.

Pour autant, l'évaluation d'expérimentations sociales reste peu développée en France, contrairement à beaucoup d'autres pays plus avancés dans ce domaine, notamment les Etats-Unis, le Royaume-Uni, le Canada ou encore l'Australie. S'il n'est pas rare en France d'expérimenter des dispositifs innovants dans le domaine social, il est plus singulier de construire des projets d'expérimentation qui intègrent un protocole d'évaluation dès leur conception. De même, s'il est fréquent de concevoir des protocoles d'évaluation, il est original qu'ils soient adaptés aux contraintes spécifiques des expérimentations sociales. Ce hiatus entre expérimentation et évaluation peut s'expliquer par le format des expérimentations qui sont souvent des micro-projets. Du point de vue de l'expérimentateur, la difficulté est de se donner un dispositif d'évaluation qui ne soit pas disproportionné par rapport au volume de l'expérimentation. Du point de vue de l'évaluateur, la difficulté est de se donner une méthodologie permettant de se prononcer

sur l'apport d'un programme innovant alors que l'échantillon des personnes qui en bénéficient est de petite taille. Il ne s'agit pas de construire une « usine à gaz » mais de se donner des moyens fiables permettant de mesurer les effets du programme. L'enjeu est de concilier des programmes sociaux innovants avec des méthodes d'évaluation rigoureuses.

Ce guide méthodologique est destiné aux porteurs de projet dans le champ des expérimentations sociales. Son objectif est de présenter de façon claire les apports et les difficultés d'une démarche expérimentale lorsqu'elle est assortie d'un dispositif d'évaluation. La présentation des différentes approches de l'évaluation se veut aussi complète que possible, compte tenu de l'état de l'art à la fin des années 2000. Ce guide est structuré autour des grandes questions qui structurent la conception d'une expérimentation et de son évaluation : Quels sont les objectifs visés par l'expérimentation ? Quels sont les enjeux de l'évaluation ? Comment mesurer au mieux les effets d'une expérimentation ? Comment expliquer ces effets ? Comment prendre en compte la perception des acteurs impliqués dans le dispositif ? Il est émaillé d'exemples d'expérimentations terminées ou en cours de réalisation.

Deux acteurs clés : Le porteur de projet et l'évaluateur

L'expérimentation est mise en œuvre par un ou plusieurs porteurs de projet (Etat, collectivité territoriale, établissement public ou association). L'expérimentation est localisée sur un terrain donné, mais a vocation à être étendue en cas de succès sur d'autres terrains et l'on s'interroge sur les conditions de cette généralisation. L'objectif peut-être par exemple de lever un obstacle à l'accès à l'emploi, de mettre en place un nouvel instrument qui favorise l'insertion sociale ou économique, ou encore de combattre un déterminant supposé de l'exclusion sociale. Cette innovation peut correspondre à un très vaste ensemble de dispositifs : programme d'accompagnement au retour à l'emploi, dispositif de lutte contre l'échec en formation initiale ou continue, lutte contre des problèmes de santé, contre l'illettrisme, etc.... Dans ces domaines et dans tous les autres, il s'agit d'améliorer un dispositif existant ou de mettre en place un nouveau programme. La question de l'évaluation doit être posée avant que l'innovation soit mise en œuvre.

L'évaluation est prise en charge par un centre de recherche ou un cabinet de conseil qui présente des références dans le champ des politiques sociales et de leur évaluation. Il maîtrise les technologies de l'évaluation adaptées à la problématique de l'expérimentation. L'évaluation est effectuée de façon externe et indépendante de l'expérimentateur. Cela n'exclut pas, bien au contraire, que l'expérimentateur mobilise des

outils internes d'observation sur le déroulement et les effets de son action. Mais ces outils ne se substituent pas à l'évaluation externe et indépendante du programme. L'évaluateur n'évalue pas l'expérimentateur ; Il évalue l'expérimentation. Par ailleurs, les résultats de l'évaluation ne sont en aucun cas donnés par avance. Cela fait de l'expérimentation et de son évaluation une activité risquée, qui implique des coûts certains pour des gains hypothétiques. En particulier, les conclusions de l'évaluation peuvent amener à renoncer à la généralisation de l'expérimentation.

Réunir les conditions de réussite d'une expérimentation sociale implique généralement une mobilisation forte des pouvoirs publics, qui doivent jouer à la fois un rôle **de maître d'ouvrage, d'intermédiaire, d'aiguillon et de facilitateur**.

En effet, ils déterminent les champs dans lesquels il y a le plus besoin d'expériences innovantes (maître d'ouvrage) ; ils favorisent la mise en relation des nombreux acteurs de l'insertion et des équipes de recherche susceptibles de les évaluer (intermédiaire) ; ils se font le garant de l'évaluation rigoureuse de l'innovation sociale (aiguillon) ; ils accordent, si besoin, les dérogations nécessaires aux règles et/ou pratiques des autorités publiques afin de pouvoir mener ces expérimentations et dégagent des fonds destinés à financer une partie de l'expérimentation (faciliteur).

Un aperçu des méthodes d'évaluation

L'évaluation mobilise un ensemble de méthodes complémentaires qui permettent de répondre à des questions différentes. Deux grands types de méthodes co-existent. On distingue les méthodes qui s'appliquent sur des échantillons représentatifs (de quelques centaines à quelques milliers d'individus, voire davantage), ou méthodes quantitatives, et les investigations approfondies sur des petits échantillons raisonnés de l'ordre de quelques dizaines d'individus, ou méthodes qualitatives. Ces deux familles de méthodes ne poursuivent pas les mêmes objets et ne répondent pas aux mêmes questions. Elles sont complémentaires. Même si les choses sont en réalité plus nuancées, on peut dire considérer globalement que les approches quantitatives répondent à la question « est-ce que le programme délivre les effets attendus ? », c'est-à-dire « est-ce que le dispositif est efficace du point de vue de l'objectif poursuivi ? ». Elles fournissent des indicateurs chiffrés de performance du programme. Les méthodes qualitatives apportent des éléments de réponses à des questions du type « comment le dispositif permet d'atteindre ces effets ? ». Elles permettent notamment de repérer les obstacles au bon fonctionnement du programme et les leviers sur lesquels il importe d'agir pour en améliorer les effets, ou de repérer les ce qui est particulièrement important lorsque l'on s'interroge sur la généralisation de l'expérimentation. Les méthodes d'évaluation ne

répondent pas aux mêmes questions et il importe donc d'être explicite sur les questions auxquelles l'évaluation doit répondre avant de faire le choix d'un protocole d'évaluation, qui combine généralement différentes méthodes.

Au sein des méthodes quantitatives, qui permettent de se prononcer sur l'efficacité du programme, il y a lieu de hiérarchiser les approches qui peuvent être retenues. Les méthodes les plus fiables, celles qui apportent la mesure la plus précise et la plus robuste de l'impact du programme sur les variables d'intérêt (outcomes), sont les méthodes d'évaluation randomisée. Elles tiennent leur appellation de l'anglais random (hasard, aléatoire). Elles reposent sur l'observation d'un groupe test et d'un groupe témoin qui ne bénéficie pas du dispositif. Les deux groupes doivent être construits en y affectant les personnes au hasard, à l'aide par exemple d'un tirage aléatoire simple dans une liste pré-établie. L'affectation aléatoire des personnes est le meilleur moyen de constituer des groupes identiques en tous points, avant l'expérimentation, ce qui permet de mesurer son effet « toutes choses égales par ailleurs ». D'autres protocoles sont envisageables dès lors qu'ils mettent en œuvre à un moment opportun un tirage au sort qui décide de l'affectation du programme entre différents groupes.

Pour mesurer les effets d'un nouveau dispositif, il faudrait idéalement pouvoir observer la situation des bénéficiaires du dispositif et la comparer avec ce qui se serait passé pour ces mêmes bénéficiaires en l'absence du dispositif. Comme cela n'est pas possible en pratique, puisqu'une même personne ne peut pas à la fois être dans le dispositif et en dehors, un point de repli consiste à construire une situation très proche de celle que l'on aurait observée si les bénéficiaires n'avaient pas eu accès au dispositif, que l'on appelle un « contre-factuel ». Le meilleur contre-factuel possible est obtenu lorsque les bénéficiaires du dispositif et les non bénéficiaires ont été tirés au hasard. On est alors assurés que toutes les caractéristiques des deux groupes sont bien les mêmes en moyenne. Dans l'ensemble, ces méthodes s'apparentent aux expériences menées en épidémiologie pour tester l'efficacité d'une nouvelle thérapie ou en marketing pour étudier l'impact d'une innovation de produit. Elles sont particulièrement bien adaptées à l'évaluation d'expérimentations sociales, dans lesquelles un groupe restreint d'individus fait l'objet d'un traitement particulier qui déroge au droit commun. Il est toujours souhaitable de privilégier une évaluation selon un protocole d'expérience randomisée dès lors que les conditions de l'expérimentation le permettent.

L'évaluation randomisée implique cependant une rupture temporaire au principe d'égalité, ce qui peut soulever dans certains cas des obstacles éthiques qui mettent en question l'opportunité de sa mise en œuvre. Si l'expérience empêche une personne d'accéder à une ressource déterminante pour sa trajectoire de vie ou son parcours professionnel, on ne peut l'envisager. Pour autant, une évaluation est le seul moyen de savoir si cette ressource est effectivement déterminante. Dans le cas le plus général, l'expérimentation sociale consiste surtout à offrir un accès privilégié à un ensemble de services offerts par l'expérimentateur. Lorsqu'il s'agit de tester les effets d'un surcroît de service, d'un supplément au droit commun, l'expérimentation ne soulève plus les mêmes obstacles éthiques. En d'autres termes, l'expérience contrôlée ne doit pas consister à priver certaines personnes de l'accès à une ressource à laquelle ils ont le droit, mais à offrir à certains un accès sélectif et temporaire à une ressource nouvelle supplémentaire afin de juger de son efficacité en vue d'en généraliser à terme le bénéfice à l'ensemble de la population éligible.

Insérer un encadré-exemple (cf comité d'éthique aux US + exemple FR)

S'il n'est pas possible de procéder à une évaluation randomisée, d'autres méthodes quantitatives peuvent être mobilisées. Elles consistent à tenter de reproduire artificiellement les conditions d'une expérience contrôlée à l'aide de techniques statistiques adaptées aux caractéristiques de l'expérimentation.

Le point de départ : définir les objectifs de l'expérimentation

La première étape dans la construction d'une expérimentation est d'identifier clairement l'objectif poursuivi. Il s'agit de savoir à la fois ce que l'on veut faire et pourquoi on veut le faire. La difficulté n'est pas seulement d'être en mesure d'explicitier un objectif, mais aussi de s'assurer qu'il est effectivement partagé par toutes les parties prenantes, par tous les acteurs qui vont contribuer à la réalisation du projet, y compris les bénéficiaires eux-mêmes. En l'absence d'un objectif clair et/ou partagé, l'expérimentation est souvent vouée à l'échec et son évaluation peut perdre toute pertinence. En outre, la définition de l'objectif conditionne le choix des variables d'intérêt qui seront retenues pour l'évaluation.

Il importe évidemment que les ressources humaines et/ou matérielles engagées dans le programme soient proportionnées aux objectifs qu'il poursuit. Cet aspect des moyens mis en œuvre ne doit être négligé ni par l'expérimentateur, ni par l'évaluateur. L'efficacité du programme et l'intérêt d'une éventuelle généralisation s'évaluent à l'aune des moyens

déployés pour sa réalisation. La première tâche de l'évaluateur est donc de prendre la mesure des objectifs et des moyens de l'expérimentation. Cela suppose d'étudier la documentation du programme et aussi de suivre sur le terrain la réalité de l'expérimentation. Même une évaluation purement quantitative ne peut faire l'économie de ce travail de terrain.

Des investigations approfondies peuvent être réalisées afin de vérifier la lisibilité de l'objectif poursuivi par l'expérimentateur. Ces approches sont de nature qualitative. Elles consistent à réaliser des enquêtes de terrain, à interroger l'ensemble des acteurs qui sont parties prenantes de l'expérimentation, à recueillir leur perception sur les objectifs et le déroulement de l'expérimentation, puis à effectuer une synthèse de toutes ces informations. Des investigations spécifiques peuvent être menées pour reconstituer la genèse de l'expérimentation : comment a germé l'idée ? Au terme de quel diagnostic et pour répondre à quel problème ? S'agit-il d'un diagnostic partagé ? etc. Dans les programmes qui sont réalisés à grande échelle, ces approches qualitatives peuvent être utilement complétées par des enquêtes statistiques qui permettent de mesurer la notoriété du programme et les attentes des bénéficiaires.

I-LES METHODES QUANTITATIVES : LA MESURE DES EFFETS

Faire le choix d'un protocole d'évaluation quantitative

Après avoir défini les objectifs visés par l'expérimentation, il importe de vérifier si ces objectifs ont été atteints en réalisant une étude d'impact. Cela consiste à mesurer les effets immédiats et différés du programme qui a été mis en œuvre sur une ou plusieurs variables d'intérêt (par exemple, le taux d'accès à l'emploi, les sorties de la pauvreté monétaire, la fréquence des échecs scolaire, etc.). Il s'agit de vérifier si les effets attendus du programme ont bien été réalisés et aussi de savoir si le programme n'a pas eu d'autres effets, qui n'étaient pas nécessairement attendus. Pour mener à bien ce type d'étude, qui constitue le cœur de l'évaluation, il est indispensable d'adopter une approche quantitative, c'est-à-dire de mobiliser des données statistiques et des moyens de traitement adaptés aux caractéristiques des données. La phase quantitative peut toutefois être précédée d'une phase qualitative, permettant par exemple de repérer dans les propos des acteurs des effets non attendus, qui seront ensuite mesurés. On peut présenter le contenu de cette étude d'impact en distinguant différentes étapes : 1) le choix d'un protocole d'évaluation ; 2) la construction d'un système d'information ; 3) le traitement des données ; 4) l'analyse des résultats, la rédaction de l'étude quantitative et de ses conclusions.

La construction d'un système d'observation et de suivi statistique de l'expérimentation est une tâche indispensable. Pour autant, elle ne se suffit pas en elle-même. Il est essentiel aussi de se donner un protocole d'évaluation qui guidera le recueil des données et qui en permettra le meilleur usage possible. L'objectif n'est pas seulement de construire des chiffres, il s'agit surtout de construire des chiffres de qualité, qui permettront de produire une mesure fiable et précise des effets de l'expérimentation.

Deux grandes familles d'approches : expériences contrôlées et « quasi-expériences »

Les spécialistes ¹ distinguent deux grandes familles de méthode d'évaluation quantitative qui se différencient par la nature des données utilisées. La première famille est celle des évaluations qui utilisent des données d'expériences contrôlées (on parle aussi d'expériences randomisées). La deuxième famille correspond aux évaluations qui mobilisent des données non expérimentales mais qui tentent de se rapprocher des conditions des données expérimentales (on parle aussi de données de quasi-expérience, ou d'expérience naturelle).

Ces deux grandes familles se hiérarchisent du point de vue de la qualité du chiffre qu'elles parviennent à produire. Les évaluations les plus précises et les plus fiables sont obtenues avec des données d'expériences contrôlées et randomisées. Aux yeux des spécialistes les plus exigeants, ces méthodes sont les seules qui permettent de véritablement prouver les effets d'une expérimentation. La qualité de la preuve est, en quelque sorte, équivalente à celle d'un flagrant-délit pour une cour de justice, alors que les autres méthodes, qui mobilisent des données non expérimentales, reviennent à accumuler des présomptions de culpabilité sans jamais véritablement administrer la preuve de façon définitive. Certes, elles réduisent les marges d'erreur, parfois de façon très significative, mais elles sont toujours moins fiables qu'une évaluation randomisée. C'est la raison pour laquelle la bonne démarche évaluative est de chercher en premier lieu à construire une expérience contrôlée et à se replier sur une méthode non contrôlée en cas d'impossibilité.

Ce qu'est une expérience randomisée

Une expérience sociale contrôlée, dans sa version la plus simple, implique de constituer deux groupes de personnes, puis à donner l'accès au dispositif que l'on souhaite évaluer

¹ En particulier le prix Nobel d'économie James Heckman, cf "Microdata, Heterogeneity and the Evaluation of Public Policy", Prize Lecture, December 8 (2000).
http://nobelprize.org/nobel_prizes/economics/laureates/2000/heckman-lecture.pdf.

à un groupe que l'on nommera le groupe test et à ne pas donner l'accès à l'autre groupe que l'on nommera le groupe témoin. Le point crucial est d'affecter au hasard les personnes dans chacun des deux groupes, en utilisant un tirage aléatoire simple dans une liste pré-constituée d'individus potentiellement éligibles au programme. Un tirage au sort est la meilleure façon pour que les personnes des deux groupes aient en moyenne les mêmes caractéristiques observables ou non observables. S'ils ont une taille suffisante (plusieurs centaines d'individus dans chaque groupe), les deux groupes auront exactement la même composition par âge, sexe, qualification, et aussi selon d'autres caractéristiques que l'on ne peut pas observer, par exemple la motivation, la capacité à coopérer avec les institutions, les opinions politiques, syndicales ou religieuses, et qui peuvent avoir un impact sur les variables d'intérêt. En outre, plus les groupes sont de grandes tailles, plus les effets observés seront précis : même de petites différences de moyennes entre les deux groupes auront une signification si les groupes sont de taille suffisante. En pratique, le degré de précision de ce que l'on souhaite observer permet de définir la taille des groupes (une étude spécifique peut être réalisée pour définir la taille des groupes).

Dans la construction du protocole, il est nécessaire de distinguer les mesures auxquelles les individus de groupe test ont accès. En effet, si ces individus bénéficient simultanément de plusieurs mesures, il sera impossible d'évaluer et de comparer leur efficacité respective. Pour tester l'efficacité de plusieurs mesures prises individuellement ou de façon combinée, il existe des protocoles d'expériences contrôlées plus sophistiqués, avec trois ou quatre groupes, avec des entrées différées dans le dispositif ou avec plusieurs sélections successives dans les groupes. Parfois, c'est le programme lui-même qui fait l'objet d'un tirage au sort, avec par exemple des traitements différenciés selon les personnes. La seule constante est qu'il est nécessaire d'avoir au moins une étape de l'expérience où un tirage au sort a lieu. Seul le hasard permet de distinguer les individus selon un processus parfaitement indépendant de leurs caractéristiques personnelles, qu'elles soient observables ou non observables.

Le problème des biais de sélection et la nécessité de produire des données contrôlées

A quoi bon une telle complexité apparente dans les méthodes d'analyse ? Pourquoi ne pas se contenter de produire des statistiques descriptives et de les commenter ? La réponse est qu'un suivi d'indicateur dans le temps n'est pas une évaluation. Par exemple, il n'est pas satisfaisant de suivre l'évolution du taux d'accès à l'emploi des bénéficiaires d'un programme d'accompagnement, même en le comparant à celui des demandeurs

d'emploi qui ne bénéficient pas du programme. De même, il n'est pas suffisant de suivre dans le temps la part de ménages pauvres parmi les bénéficiaires d'un dispositif expérimental de soutien aux bas revenus, même si on la compare aux ménages qui ne bénéficient pas du programme.

La raison est simple : les personnes qui ont bénéficié du programme social n'ont pas forcément les mêmes caractéristiques que celles qui n'en ont pas bénéficié. Les bénéficiaires d'un programme social ont été sélectionnés sur la base d'un ensemble de caractéristiques qui ont des effets sur les trajectoires personnelles. Si l'on ne tient pas compte de ces différences entre la population qui bénéficie du programme et celle qui n'en bénéficie pas, on fausse l'évaluation des résultats du traitement. Pour qualifier ce phénomène, on parle de « biais de sélection ». Le problème est lié au fait que l'échantillon des personnes qui vont bénéficier du programme social n'est pas représentatif de l'ensemble des éligibles. Il y a eu une erreur systématique lors de la sélection des personnes qui vont bénéficier du programme.

Les caractéristiques qui influencent les trajectoires et qui peuvent contribuer à la présence d'un biais de sélection sont de nature variées. Certaines sont observables. Par exemple, il existe de nombreuses études qui montrent que les chances de trouver un emploi lorsque l'on est au chômage baissent avec l'âge, augmentent avec la qualification, sont plus faibles pour les femmes, etc. Si le groupe de bénéficiaires comprend plus d'hommes, de jeunes, et de qualifiés il va présenter un taux de retour à l'emploi mécaniquement plus élevé que celui des non bénéficiaires. Dans ce cas, le risque est alors de surévaluer les effets du programme. Si celui-ci est ensuite généralisé à l'ensemble de la population éligible, il ne produira pas les effets attendus. Ce type de biais sur les caractéristiques peut être convenablement corrigé avec des méthodes économétriques adaptées. Mais d'autres caractéristiques ne sont pas observables. Par exemple, la motivation, la carrière professionnelle antérieure, les accidents de vie, les loisirs et les pratiques culturelles, etc. peuvent exercer un effet sur les chances d'insertion mais constituent autant de variables qui ne peuvent pas être observées par l'évaluateur. Ces variables peuvent contribuer à un biais de sélection sans que l'on sache a priori dans quel sens joue ce biais.

Pour produire une mesure fiable des effets d'un programme, il est donc indispensable de s'affranchir de ces différents biais. Les spécialistes disent qu'il faut contrôler de l'hétérogénéité des variables observables et des inobservables. Pour y parvenir, il est nécessaire de mobiliser des techniques très sophistiquées qui requièrent un grand nombre d'observations, ou de construire des protocoles permettant de contrôler les

données. L'intérêt majeur d'une expérimentation sociale qui intègre une évaluation selon un protocole randomisé, est de s'affranchir de ces biais de sélection sur les individus bénéficiant du programme de façon à produire un chiffrage très robuste avec une grande économie de moyen statistique ou économétrique.

Pour autant, une expérimentation randomisée peut elle-même être confrontée à l'existence de biais de sélection dans le choix des terrains d'expérimentations et des institutions expérimentatrices. 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. Ainsi, à l'origine d'une expérimentation se trouve une équipe d'expérimentateurs potentiels porteurs d'un projet. Par nature, ce type d'institution est favorable à ce type de protocole et est prêt à fournir les efforts nécessaires à sa mise en œuvre. Les effets de la mesure testée sur le groupe test peuvent s'en trouver influencés. Une façon de s'affranchir de ce biais de sélection consiste à conduire simultanément l'expérimentation sur des sites différents, avec des expérimentateurs différents. La sélection des sites s'effectuant si possible par tirage au sort. Les mêmes mesures seront alors testées en mobilisant les mêmes instruments et en suivant les mêmes protocoles, de sorte que toute différence d'effet de ces mesures entre les sites sera imputable aux spécificités de ces derniers.

Les obstacles éthiques

Les limites d'une approche randomisée méritent aussi d'être abordées. Même si le tirage au sort est le meilleur moyen de garantir ex ante l'égalité des chances, le fait de séparer la population éligible en un groupe test et un groupe témoin se traduit par une inégalité ex post de situation. Certes, cette rupture au principe d'égalité se justifie par le caractère temporaire de l'expérimentation et par la perspective de sa généralisation ultérieure en cas de succès. Mais elle n'en reste pas moins contraire au principe d'égalité de traitement des personnes, ce qui pose une question éthique. De surcroît, un protocole d'expérimentation randomisée implique de priver une partie de la population des ressources qui pourraient lui être nécessaires pour améliorer sa situation. 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 dont la finalité première est d'améliorer le bien-être de personnes en difficulté d'insertion sociale et/ou économique.

La question éthique ne doit pas être occultée. Elle est souvent soulevée par les travailleurs sociaux qui sont en contact direct avec les publics traités ou encore par les

élus qui souhaitent naturellement que le plus grand nombre de bénéficiaires aient un accès immédiat à l'innovation sociale. Il est souhaitable qu'elle soit prise en considération dans toutes les expérimentations sociales, au cas par cas, quel que soit leur protocole d'évaluation, et plus encore lorsqu'est envisagée une évaluation randomisée. Parfois même, cette réflexion pourra conduire à abandonner l'idée d'une évaluation par tirage au sort. C'est le cas lorsque les expérimentations mettent en jeu des ressources vitales ou déterminantes pour la trajectoire de vie des personnes. Le problème se pose notamment dans de nombreux programmes d'aides mis en œuvre dans les pays en développement². Il n'est pas question d'organiser à des fins expérimentales une privation de ressource vitale pour des personnes en difficulté d'insertion sociale. Pour cette raison, il s'agit d'appliquer les méthodes d'évaluations randomisées uniquement sur des expérimentations où l'on veut tester un « plus » par rapport au droit commun. Le groupe témoin dispose de l'accès aux ressources de droit commun. Le groupe test dispose de surcroît d'un accès à un complément de ressource, à un service amélioré, dont il s'agit de tester l'efficacité.

Le tirage au sort peut donc être considéré comme un moyen plus juste, plus transparent et plus légitime que l'affectation arbitraire d'une ressource rare au sein d'une population éligible. Dans tous les cas, l'accès au programme est généralement limité par les moyens humains ou financiers consacrés à l'expérimentation, la randomisation est alors une solution pour donner un accès sélectif au programme. Les situations de rationnement imposent de sélectionner les personnes qui ont accès au programme ou d'organiser une graduation d'accès dans le temps et le tirage au sort est un bon moyen d'organiser cette sélection ou cet accès différé. Il est d'ailleurs parfaitement compatible avec l'affectation selon une sélection raisonnée, en mobilisant des critères sociaux visant à prioriser des publics cibles. Pour combiner les deux modes d'affectation, il suffit en effet de définir une liste de personnes éligibles au programme en mobilisant ces critères, puis d'effectuer un tirage au sort au sein de cette liste de façon à constituer le groupe test et le groupe témoin.

Il est en outre très souvent possible en pratique de modifier les paramètres de l'expérimentation et de son évaluation, afin de lever les obstacles à la mise en œuvre d'un protocole randomisé. Par exemple, si l'on ne peut pas priver certaines personnes de

² Pour une synthèse sur ces évaluations appliquées aux pays en développement, voir Abhijit V. Banerjee et Esther Duflo, « The Experimental Approach to Development Economics », miméo MIT, <http://econ-www.mit.edu/files/3158>

l'accès au programme social, on peut envisager de le reporter dans le temps. Le groupe test est constitué des personnes qui ont un accès immédiat au programme. Le groupe témoin est constitué de ceux qui auront un accès ultérieur, par exemple six mois plus tard. Si des différences sont observées entre les deux groupes du point de vue des variables d'intérêt, elles pourront être attribuées au fait d'être effectivement passé par le programme. Ce type de protocole garantit que tous les éligibles auront effectivement accès au dispositif, ce qui est souvent une attente légitime de l'expérimentateur.

Par ailleurs, si le tirage au sort ne peut être effectué sur les personnes, il est également possible de randomiser le traitement. Par exemple, dans un protocole où l'on évalue l'accès à des dispositifs de micro-crédits dans des villes moyennes, où il n'est pas possible d'affecter au hasard la possibilité de bénéficier du programme, il est envisageable de tenter de moduler de façon aléatoire les conditions du crédit, son coût, sa durée, les conditions de garanties, etc. ce qui permettra in fine de réaliser une évaluation satisfaisante sur le plan de la rigueur scientifique.

Ces exemples illustrent le fait que la prise en compte des questions éthiques n'est pas incompatible avec le montage d'évaluation randomisée même si elle impose un effort de réflexion commune à l'expérimentateur et à l'évaluateur. Il s'agit bien souvent de trouver des solutions astucieuses à des questionnements et à des situations toujours particulières. Il n'en reste pas moins que tous les programmes sociaux ne sont pas évaluables selon une comparaison entre un groupe test et un groupe témoin avec assignation aléatoire des personnes dans les deux groupes.

Un aperçu des approches quantitatives alternatives

S'il est impossible de conduire une expérimentation sociale qui intègre une évaluation selon un protocole randomisé, on peut avoir recourt à une évaluation sur des données non expérimentales. Une simple comparaison des moyennes des variables de performance des individus bénéficiant de la mesure et de ceux qui n'est bénéficiant pas serait toutefois incorrecte, du fait de l'existence des biais de sélection. Il convient alors d'appliquer des techniques statistiques et économétriques permettant de corriger au mieux ces biais³.

Idéalement, il conviendrait de comparer la variable de performance d'un même individu selon qu'il bénéficie ou non d'une mesure ou encore d'un programme, dont l'efficacité est

³ Pour une présentation complète de ces techniques, voir Angrist J. et A. Krueger, (1999), Dehejia R. et Wahba S., (1999) ou Meyer B. (1995).

à tester. Or, par définition, soit un individu donné en bénéficiant, soit il n'en bénéficie pas. Dans l'un ou l'autre cas, on observe donc une seule valeur de la variable de performance pour chaque individu. Il faut donc estimer la variable de performance d'un individu bénéficiant de la mesure s'il n'en bénéficiait pas, et la variable de performance d'un individu ne bénéficiant pas de la mesure s'il en bénéficiait. Pour y parvenir, les experts mobilisent une approche qu'ils qualifient de « méthode d'appariement »⁴. L'intuition est la suivante. Si deux individus ont la même probabilité de bénéficier de la mesure, et que le premier l'est alors que le second ne l'est pas, tout se passe comme si l'accès à la mesure était affecté aléatoirement entre ces deux individus. Dans ce cas, la comparaison des variables de performance de ces deux individus fournirait une estimation de l'effet de la mesure sur la variable de performance. Pour estimer l'effet de la mesure, on considère les individus qui en bénéficient et, pour chacun d'entre eux, on estime ce que serait leur variable de performance en l'absence de la mesure. Pour cela, on sélectionne un individu bénéficiant de la mesure et on effectue la moyenne des variables de performance des individus n'en bénéficiant pas qui ont la même probabilité que lui d'en bénéficier. On effectue ensuite la moyenne des écarts obtenus sur tous les individus bénéficiant de la mesure. On effectue une opération symétrique pour estimer l'effet potentiel qu'aurait la mesure sur les individus n'en bénéficiant pas, et on estime l'effet total en appariant chaque individu avec ceux du groupe auquel il n'appartient pas.

Le principe de cette méthode consiste donc à comparer la variable de performance de chaque individu bénéficiant de la mesure, avec celle d'un individu « clone » ne bénéficiant pas de la mesure. On considère que ces deux individus sont « clones », autrement dit qu'ils ont les mêmes caractéristiques dans la mesure où ils ont la même probabilité de bénéficier de la mesure. Apparaît ici une limite importante de cette méthode. Si un petit nombre de caractéristiques influencent la probabilité de bénéficier de la mesure, on considérera comme « clones » deux individus qui dans les faits auront un grand nombre de caractéristiques distinctes, et de fait, on attribuera à la mesure des effets qui seront en partie liés à ces différences de caractéristiques non contrôlées.

On le voit, les méthodes alternatives à l'évaluation randomisée sont très sophistiquées sur le plan des techniques statistiques. Elles requièrent d'être mise en œuvre par des experts confirmés et ne garantissent pas l'absence de tout biais. Surtout, elles supposent pour être mises en œuvre de pouvoir disposer d'un grand nombre de d'observations, ce qui n'est pas toujours compatible avec une expérimentation sociale qui porte le plus souvent sur des échantillons de quelques dizaines ou quelques centaines de bénéficiaires. Pour mener à bien une évaluation quantitative permettant de produire une mesure fiable des

⁴ Cette méthode a été proposée initialement dans le champ de l'économie de l'éducation par D. Rubin (1974).

effets d'un programme social, il n'y a donc guère d'alternative au choix d'un protocole d'évaluation randomisée lorsque l'expérimentation est de petite taille.

Construire et mobiliser un système d'observation

En pratique, une étape préalable de l'évaluation consiste à expertiser le système d'information qui va permettre de suivre les variables d'intérêt dans le temps. Il faut se donner des outils d'observation, construire des indicateurs, recueillir des données et déployer des technologies de traitement pour ces données. Le système d'observation pourra le plus souvent être constitué des fichiers informatiques de gestion utilisés de façon régulière par l'expérimentateur. Mais dans la plupart des cas, il sera nécessaire de compléter ces fichiers par des données spécifiques à l'expérimentation. Par exemple, on aura besoin de savoir pour chaque individu sa date d'entrée dans le programme, la durée de séjour dans le dispositif, et des éléments sur sa situation après la sortie du programme. On aura besoin de compléter les informations contenues dans le fichier de gestion par des enquêtes statistiques réalisées auprès de bénéficiaires ou de non bénéficiaires de l'expérimentation. Eventuellement, il sera nécessaire de répéter ces enquêtes complémentaires plusieurs fois afin de suivre les populations dans le temps. Pour recueillir ce type de données, il est nécessaire de réaliser un travail spécifique d'évaluation du système d'information de l'expérimentateur et des moyens de le compléter. Ce travail fait partie intégrante de l'évaluation et sera réalisé par l'évaluateur qui doit ainsi procéder à une sorte d'audit préalable des systèmes d'information.

L'évaluation quantitative requiert la constitution de bases de données ayant à la fois une dimension individuelle et temporelle. Ces bases agrègent deux types de sources statistiques. D'un côté, on mobilise les bordereaux de suivi de l'expérimentation, qui consistent dans des fiches individuelles qui donnent la date d'entrée dans le programme, le type de parcours suivi dans le programme, la date de sortie, et d'autres indications sur le déroulement du programme et la situation des personnes avant et après le programme, ainsi que des fiches collectives, sur les différents opérateurs et leurs prestations. D'un autre côté, on mobilise des fichiers administratifs de gestion qui permettent de suivre les personnes du point de vue de l'expérimentateur et des organismes gestionnaires (tels que Pôle Emploi ou la CAF). Toutes ces sources statistiques doivent être appariées de façon à disposer d'une information complète sur les trajectoires et les caractéristiques des personnes avant, pendant et après l'expérimentation, qu'elles appartiennent au groupe test ou au groupe témoin. Puis, des traitements statistiques et économétriques sont réalisés sur ces bases de données.

La production d'une telle base de données est un travail statistique lourd. S'agissant de données individuelles, elle peut poser en outre des questions de secret statistique et de respect du cadre législatif sur l'informatique et les libertés. L'expérimentateur est le producteur et le gestionnaire des fichiers de gestion. Il organise également, de façon directe ou indirecte, la production des données de suivis tout au long de l'expérimentation, avec l'appui technique de l'évaluateur. Ce dernier est destinataire de la base donnée. Afin de respecter les textes en vigueur relatifs au secret statistique et à la confidentialité des données individuelles, l'évaluateur ne doit disposer ni même manipuler aucune information directement ou indirectement nominative. La base doit donc être complètement anonyme et les appariements de fichiers doivent être effectués en amont, par l'expérimentateur. Il s'agit là encore d'un surcroît de travail non négligeable. Cet investissement pourra cependant être amorti à la fois à des fins d'évaluation interne et externe, à des fins descriptive et explicative. La constitution de ce système d'observation pourra par ailleurs permettre d'améliorer la gestion même du dispositif.

II- LES METHODES QUALITATIVES : QUELS EFFETS ET POURQUOI CES EFFETS ?

Pourquoi une approche qualitative dans un processus d'évaluation ?

Si les méthodes quantitatives ont un fort pouvoir d'objectivation, et sont indispensables pour une approche comparative, elles ne peuvent pas toujours prendre en compte tous les aspects de l'évaluation.

L'approche qualitative peut alors être utilisée soit en combinaison avec une approche quantitative, soit dans une démarche d'investigation spécifique.

En combinaison avec une approche quantitative, les méthodes qualitatives peuvent intervenir soit en amont, soit en aval :

- En amont, elles contribuent à tester des hypothèses (et notamment repérer quels sont les effets qui devront ensuite être mesurés) et des choix de variables, et à identifier les informations disponibles. Cette phase exploratoire est particulièrement importante en situation expérimentale, où, par définition, les dispositifs sont innovants et peuvent donc générer des comportements inattendus.
- En aval, elles visent à formuler des types d'interprétation des données obtenues, à valider des hypothèses explicatives, à donner un cadre global d'analyse permettant d'interroger la cohérence des résultats observés. Cette phase

compréhensive a du sens dans les dynamiques expérimentales car on a parfois des difficultés à comprendre certaines corrélations.

Elles permettent enfin de compléter les résultats quantitatifs en apportant des éléments de réponses sur le fonctionnement d'un dispositif. Une des principales difficultés de cette analyse du fonctionnement est de faire la part des choses entre ce qui relève du caractère expérimental et ce qui relève du dispositif en lui-même. Par exemple, sur la question des coûts, si le coût de versement d'une prestation est a priori homothétique entre le dispositif expérimental et le dispositif généralisé, il y a fort à parier que les coûts de gestion seront moindres : une expérimentation nécessite souvent une gestion manuelle, alors que les processus sont automatisés en mode généralisé. Mais l'économie réalisée est difficile à chiffrer.

Les protocoles d'évaluation portant sur des politiques complexes associent fréquemment la mise en œuvre de démarches quantitatives et qualitatives, ce qui semble nécessaire pour appréhender globalement leurs conditions de fonctionnement et les résultats obtenus.

Mais l'approche qualitative se justifie aussi comme démarche spécifique appropriée à certains types de questionnement qui ne peuvent pas, ou difficilement, faire l'objet de mesures quantitatives. Face à l'évaluation objectivée par la mesure, elle introduit un jugement de type inductif et compréhensif.

Les démarches qualitatives peuvent apporter des informations sur différents types de questionnement.

- Elles peuvent permettre d'abord la mise à jour des conceptions et représentations sociales des différentes catégories d'acteurs concernés, en incluant parmi eux les bénéficiaires ou usagers des politiques étudiées. Ces représentations souvent implicites déterminent les conduites, comportements, attitudes et opinions des différents groupes d'acteurs et leur manière de s'approprier les mesures et dispositifs analysés (appropriation par les bénéficiaires notamment). Elles permettent de rendre compte de la diversité des attentes et des besoins.

- Par des démarches qualitatives visant à créer des situations d'interactivité, on peut par ailleurs favoriser une appréhension globale des situations et chercher à dégager des consensus sur des questions complexes. Le jugement évaluatif est fonction des valeurs de référence portées par les différents acteurs, et la confrontation des systèmes de valeur en présence permet de former un jugement sur la pertinence et la validité des politiques

mises en œuvre à partir de critères partagés. Cette approche pluraliste est une garantie apportée contre les risques de jugements évaluatifs normatifs liés à la seule mise en relation de résultats et d'objectifs.

Aussi ces démarches sont-elles nécessairement de nature participative.

- Les méthodes qualitatives peuvent également permettre de mesurer la diversité des situations, en s'appuyant, par exemple, sur l'analyse de cas extrêmes prenant valeur de cas types. Elles permettent aussi d'observer le déroulement de processus dans le temps, sur la base d'analyses de cas et de monographies, et de prendre en compte dans la durée des dysfonctionnements et des effets non prévus ou secondaires. Aussi les évaluations qualitatives sont-elles souvent menées en continu, de manière parallèle à l'action, sur longue durée.

Inductive, compréhensive, participative, en continu, telles sont les caractéristiques fréquemment associées d'une évaluation qualitative. Les enjeux multiples auxquels elle doit répondre nécessitent le plus souvent la combinaison de plusieurs approches.

Quelques méthodes d'évaluation qualitative

Les méthodes et techniques d'évaluation qualitative doivent être choisies d'abord en fonction des questions que l'on se pose, de ce qui est attendu principalement comme résultat évaluatif.

Ces finalités étant clairement posées, les techniques elles-mêmes sont largement empruntées aux méthodes des sciences sociales, et ne sont pas très spécifiques.

Les investigations qualitatives peuvent prendre trois grandes formes : l'interrogation des acteurs, l'observation directe de l'évaluateur, la monographie.

- L'interrogation des acteurs utilise toutes les techniques d'entretien approfondi, le plus souvent des entretiens semi-directifs ou ouverts.

Deux grandes formes sont utilisées, selon les objectifs recherchés, et combinées le plus souvent : entretiens individuels ou entretiens collectifs. Ces entretiens peuvent être menés auprès de différents types d'acteurs, et de manière privilégiée auprès de bénéficiaires ou d'utilisateurs.

Exemple d'utilisation d'entretiens individuels : un projet visant à mesurer les effets sociaux d'une distribution de légumes et fruits frais dans l'aide alimentaire a prévu de combiner évaluation quantitative par questionnaires, et évaluation qualitative par entretiens individuels semi-directifs. Ces entretiens avaient pour objectif d'identifier les modifications de comportement, les obstacles et difficultés rencontrés et les effets induits de cette consommation sur les liens sociaux.

Les entretiens collectifs permettent, en jouant sur les interactions, d'obtenir de nombreuses informations en un temps réduit. Ils peuvent aussi, dans la méthode dite des « focus groupes », faire émerger progressivement, par interrogations successives, des points de consensus.

Exemple d'utilisation de focus groupe: un projet visant à articuler aide alimentaire et action sociale a bâti son processus évaluatif sur des entretiens collectifs successifs auprès des trois collèges d'acteurs : les bénéficiaires, les acteurs professionnels, les institutions concepteurs du projet. Les résultats sont obtenus par la confrontation entre ces entretiens.

- L'observation directe s'inspire des méthodes de type ethnologique. L'évaluateur s'immerge dans l'organisation, selon des modalités et à des moments débattus avec les acteurs, en position d'observateur sans intervention directe. C'est une approche entièrement inductive.

Exemple d'observation directe : un projet visant à promouvoir des innovations socio-éducatives dans la protection de l'enfance. L'objectif est de conceptualiser les innovations et d'en dresser une typologie. L'immersion de l'évaluateur dans les dispositifs évalués vise à permettre une compréhension des pratiques d'intervention, des méthodes pédagogiques, des changements de comportement.

- La monographie ou l'étude de cas permet d'analyser dans la durée et de façon globale les effets d'une politique, et est particulièrement appropriée pour étudier des programmes territorialisés, avec un fort impact de l'environnement institutionnel. La comparaison entre monographies réalisées dans des contextes très divers peut aider à isoler le ou les facteurs explicatifs déterminants. L'étude de cas est aussi bien adaptée à l'analyse des fonctionnements et des processus.

Exemples d'évaluation par monographie : un projet visant à développer des programmes territorialisés et renforcés de formation et d'accompagnement vers l'emploi. A côté d'une évaluation quantitative portant sur les effets du programme sur des parcours individuels, il est prévu une comparaison territoriale pour apprécier la construction du dispositif, le jeu des acteurs, les effets collectifs.

Autre exemple : un projet visant à développer l'intégration des structures d'insertion par l'économique dans un développement local. L'évaluation, participative et en continu, sur 8 terrains d'action, procède également d'une approche par confrontation de monographies territoriales.

Ces méthodes qualitatives sont souvent combinées entre elles.

Elles doivent être choisies en fonction des objectifs de l'évaluation, soigneusement déterminés et doivent faire l'objet d'une grande rigueur dans leur mise en œuvre, selon un protocole écrit, partagé, et avec des outils (questionnaires, grilles d'observation...) soigneusement élaborés, et soumis à validation auprès d'une instance d'évaluation. Dans ces approches faisant davantage appel à la subjectivité de l'évaluateur, le contrôle par cette instance est d'autant plus indispensable. L'interprétation des résultats doit aussi faire l'objet d'une mise en débat entre toutes les parties prenantes.

Une méthode d'analyse du partenariat

Le partenariat est devenu une méthode incontournable de travail dans les politiques sociales. Son bon fonctionnement est donc un des éléments importants dans la réussite des dispositifs.

Une méthode classique d'analyse du partenariat est l'interrogation de l'ensemble des partenaires et l'analyse de leur discours autour des points de convergence. Une méthode alternative est de réaliser également une interrogation, mais de faire ensuite une analyse fondée sur les convergences ET sur les divergences. Cette seconde option permet de repérer les points de discordance dans l'analyse des faits, des postures institutionnelles et des opinions sur le fonctionnement (trois volets sur lesquels les interrogations doivent porter).

On repère ainsi dans une étude récente sur l'offre de service petite enfance (dispositif expérimental mis en place par près de 50 Caf) que les partenaires d'un même territoire ne sont pas toujours d'accord sur l'existence même d'un schéma départemental de la petite enfance, sur l'institution pilote de la Cdaje, sur le nombre de commissions, les axes dominants des travaux la qualité du partenariat ... Cette incohérence peut impacter la qualité du programme mis en œuvre.

IV Les obstacles à la mise en œuvre d'expérimentation

L'expérimentation sociale n'a pas vocation à s'appliquer à toutes les situations. Elle ne semble notamment pas pertinente : face à une situation d'urgence financière ou sociale ; quand les éléments pour la décision sont insuffisants ; quand le débat est insuffisant sur le sens des mesures à expérimenter ou quand les clivages existants sur les résultats à attendre de ces mesures ne sont pas clarifiés.

Parce qu'une expérimentation est toujours innovante, elle bouscule les routines organisationnelles, les façons de faire, les pratiques habituelles des institutions. Dès lors,

la réalisation d'une innovation sociale confronte bien souvent l'expérimentateur à de multiples sources de rigidités insoupçonnées. Il avance en terra incognita, et même s'il a réalisé un travail prospectif très complet en amont du lancement de son programme, il peut s'attendre de façon certaine à rencontrer beaucoup d'imprévus et à se heurter à de multiples obstacles, petits et grands.

Ces obstacles seront d'ordre juridiques ou réglementaires (en relation avec la multiplicité des acteurs et des contrats ou conventions qui vont les relier), d'ordre purement logistiques (les conditions d'accueil des personnes vont soulever de nombreux problèmes de disponibilité, de sécurité des locaux, ...) ou encore, liés aux chaînes de traitement des données de gestion et aux systèmes d'information (qui peuvent s'avérer incomplets ou inadaptés). Il en va de même, dans une moindre mesure, pour l'évaluateur qui 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.

Dans ce contexte, il faut souligner que le fait de vouloir mener à bien une évaluation rigoureuse des effets d'un programme expérimental ajoute des complications et des incertitudes dans le déroulement d'une expérimentation. Il en résulte des coûts de mise en œuvre, de suivi et de coordination pour les parties prenantes. Ces coûts ne sont pas forcément déterminants dans le choix stratégique du protocole d'évaluation mais ils doivent néanmoins être pris en considération par l'expérimentateur et son évaluateur.

Références

Angrist J. et A. Krueger, (1999), « Empirical Strategies in Labor Economics », in Handbook of Labor Economics, vol 3A, Ashenfelter O. et Card D. (eds.), North Holland, Amsterdam, p. 1277-1366.

Dehejia R. et Wahba S., (1999), « Causal Effects in Non experimental Studies: re-evaluating the evaluation of training programs », Journal of the American Statistical Association, 94, p. 1053- 1062.

Banerjee, Abhijit et Duflo, Esther (2008), The experimental Approach to Development Economics, Mimeo MIT.

Heckman, J. (2000). "Microdata, Heterogeneity and the Evaluation of Public Policy", Prize Lecture, December 8.

http://nobelprize.org/nobel_prizes/economics/laureates/2000/heckman-lecture.pdf

Heckman J.J., LaLonde R.J. et Smith J.A., (2000), "The Economics and Econometrics of Active Labor Market Programs", in Handbook of Labor Economics, vol 3A, Ashenfelter O. et Card D. (eds.), North Holland, Amsterdam, p. 1865-2097.

Meyer B., (1995), « Natural and Quasi-Experiments in Economics », Journal of Business and Economic Statistics, 13(2), p. 151-62.

Rubin D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies", Journal of Educational Psychology, (66): p. 688–701.