

International Initiative for Impact Evaluation



WORKING PAPER 4
IN FRENCH

Méthodologie de l'évaluation d'impact : présentation de différentes approches

Robert Chambers, Dean Karlan, Martin Ravallion and Patricia
Rogers
Juillet 2009

À propos de 3ie

L'Initiative internationale pour l'évaluation d'impact (3ie) œuvre à l'amélioration de la vie des populations du monde en développement en soutenant la production et l'utilisation de données probantes sur ce qui marche, quand, pourquoi et à quel prix. 3ie est une nouvelle initiative qui répond à la demande de données de plus grande qualité et qui renforcera l'efficacité du développement en encourageant des politiques mieux informées. 3ie finance des évaluations d'impact de qualité et milite pour éclairer la conception des programmes et des politiques dans les pays en développement.

La série Documents de travail de 3ie couvre à la fois des questions conceptuelles relatives à l'évaluation d'impact et des constats d'études spécifiques ou de revues de synthèse.

Ce document de travail est publié sous la direction de Howard White, Directeur exécutif de 3ie.

© 3ie, 2009

Contacts

Initiative internationale pour l'évaluation d'impact
c/o Global Development Network
Post Box No. 7510
Vasant Kunj P.O.
New Delhi – 110070, Inde
Tél. : +91-11-2613-9494/6885
www.3ieimpact.org

Sommaire

Préface

Howard White, Directeur exécutif, Initiative internationale pour l'évaluation d'impact (3ie)

Pour que les pauvres comptent davantage : les méthodes participatives au service de l'évaluation d'impact

Robert Chambers, Institute of Development Studies, Université du Sussex

Quelques réflexions sur les essais randomisés pour l'évaluation du développement : présentation à la conférence du Caire sur l'évaluation

Dean Karlan, Université de Yale et Innovations for Poverty Action, affiliée au Jameel Poverty Action Lab

Évaluation de trois interventions stylisées

Martin Ravallion

Banque mondiale

Pour une adaptation du protocole de l'évaluation d'impact à la nature de l'intervention et à l'objet de l'évaluation

Patricia Rogers, Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation, Royal Melbourne Institute of Technology

Préface

Les débats sur les approches conceptuelles de l'évaluation d'impact semblent depuis quelques années dans l'impasse. Face à ce blocage, 3ie, NONIE, AfrEA et l'UNICEF ont organisé au Caire du 29 mars au 2 avril 2009 une conférence internationale intitulée *Perspectives on Impact Evaluation*, qui avait notamment pour ambition de confronter différents points de vue pour faire émerger un consensus. L'une des sessions plénières a joué un rôle pivot dans cette démarche. Partant du constat que les débats s'enlisent lorsqu'ils se cantonnent au niveau conceptuel, mais qu'un plus grand consensus peut être dégagé dès qu'on en vient aux caractéristiques précises du protocole d'une évaluation particulière, il a été demandé à des experts de différents horizons d'exposer la méthodologie qu'ils retiendraient pour évaluer trois interventions fictives – une allocation sociale conditionnelle, un projet d'infrastructure et un programme de lutte contre la corruption. Je suis très heureux que les quatre intervenants aient accepté de consigner ici leurs points de vue afin d'élargir leur diffusion.

Je remercie Hugh Waddington et Rizwana Siddiqui pour leur assistance à la préparation de ce recueil.

Howard White
Directeur exécutif, 3ie

Pour que les pauvres comptent davantage : les méthodes participatives au service de l'évaluation d'impact

*Robert Chambers, Institute of Development Studies, Université du Sussex**

Résumé

Le point de départ de toute évaluation consiste à se demander pourquoi on l'effectue, qui en bénéficiera, quel impact elle aura et comment. Les approches et méthodes participatives s'inscrivent dans un paradigme pluraliste, évolutif et itératif ; ce sont l'analyse des parties prenantes, le recueil de témoignages oraux, la cartographie sociale participative, l'établissement de graphiques et diagrammes de relations causales, de tendances et de changements, le calcul de scores et le brainstorming sur les forces et les faiblesses des programmes. Bien conçues et mises en œuvre, les méthodes participatives sont rigoureuses et apportent des éclairages qualitatifs, mais elles peuvent aussi quantifier le non-quantifiable et produire des statistiques sur des dimensions pertinentes qui passeraient autrement inaperçues ou seraient jugées exclusivement qualitatives. Elles ouvrent les études aux points de vue des individus les plus touchés par un projet par des moyens inaccessibles aux méthodes traditionnelles et permettent de donner plus de poids à la réalité et à l'expérience des pauvres.

Remarques liminaires

Bien que cette invitation m'honore, ma relative incompetence m'a fait hésiter à l'accepter. En effet, la majorité des personnes ici présentes [plus de la moitié d'après le vote à main levée] ont conçu des évaluations. Je n'en ai conçu qu'une ; c'était dans les années 70 et ce fut un désastre. L'autre raison pour laquelle j'ai hésité, c'est qu'on assiste aujourd'hui à une véritable explosion des innovations dans le domaine des méthodologies participatives, notamment pour le suivi et l'évaluation, et que les trois programmes choisis se prêtaient moins à ces méthodes que ne l'auraient fait des programmes axés, par exemple, sur le développement communautaire, l'agriculture ou la gestion des ressources naturelles. Enfin, je suis gêné, comme d'autres conférenciers l'ont été, de venir grossir les rangs des intervenants du Nord. Pour reprendre les termes d'un Africain lors d'une session non plénière : « tout ce qui vient du Nord a valeur de vérité ».

Reconnaissons qu'une grande part de la « vérité » – la créativité et l'innovation – en matière de méthodologies participatives est venue et vient du Sud, de l'Asie, de l'Amérique latine et, particulièrement, de l'Afrique.

J'ai néanmoins été encouragé par l'insistance de Sulley Gariba sur « l'autonomisation des communautés » dans son introduction et par l'appel d'Erma Manoncourt à « ouvrir la porte à la participation et à l'autonomie des peuples ».

Dans les trois cas considérés ici – allocation sociale conditionnelle, développement d'infrastructures et commission anticorruption, je plaiderais pour une approche pluraliste, évolutive et itérative. Des méthodes mixtes seraient appliquées. Pour commencer, il faudrait s'interroger sur l'économie politique de l'évaluation : qui en bénéficierait ? Qui pourrait y perdre ? Et comment ? Et surtout, quelles suites concrètes il serait prévu de donner aux constats. Cela pourrait nécessiter un atelier de brainstorming avec du personnel de l'agence bailleur de fonds. Si celle-ci n'était pas disposée à y consacrer du temps ou à éclairer ces questions, ou si elle n'en avait pas la capacité, j'espère que j'aurais le courage et les moyens de refuser la mission. Je

* Contact : robertc@ids.ac.uk

négocierais pour que le protocole d'accord prévoie d'autres étapes, en particulier l'analyse des parties prenantes et des négociations pour associer les parties prenantes concernées au processus.

Je limiterai essentiellement mes suggestions aux méthodes participatives qui pourraient remplacer ou compléter d'autres méthodes. Pour celles-ci, la priorité serait de rechercher de bons facilitateurs-innovateurs. Je négocierais des délais (sans doute plus longs que ceux prévus par les bailleurs de fonds) pour les ateliers et le travail sur le terrain afin de concevoir et de tester un projet pilote de démarche participative, en évitant d'y mettre fin prématurément et en restant candide, ou pour reprendre le titre de la thèse de doctorat d'Irene Guijt, en « recherchant l'étonnement¹ ». Mais surtout, et tout au long du processus, il y aurait la question de l'utilité de l'impact qu'aurait l'évaluation d'impact elle-même, de son rapport coût-efficacité et de sa valeur ajoutée.

Cas n° 1 : allocation sociale conditionnelle

Les méthodes et démarches participatives suivantes seraient envisagées :

- Cartographie participative de recensement des communautés représentatives. Il serait important de soigneusement sélectionner les communautés. Le recensement couvrirait tous les membres de chaque communauté afin d'éviter les problèmes d'échantillonnage et viserait à dénombrer les filles d'âge scolaire, scolarisées et non scolarisées. Cette opération produirait des statistiques à croiser avec les registres officiels, qui couvriraient les modifications des performances des filles.
- Études de cas individuels, comprenant des personnes extérieures et des mécontents
- Focus groupes et entretiens semi-structurés en fonction des besoins avec des filles, des enseignants, des parents, des administrateurs ...
- Invitation donnée à des filles, mais aussi à d'autres, de créer et de jouer une représentation de leur vie et de leur expérience avant et après
- Facilitation de l'établissement de diagrammes de causalité
- Recueil de témoignages
- Brainstorming sur les forces et les faiblesses du programme et sur les améliorations possibles.

Les facilitateurs/chercheurs pourraient séjourner quelques jours dans les communautés et se réunir pour comparer leurs notes. Cet exercice ferait certainement apparaître de nouveaux problèmes et questions, qui viendraient grossir le programme de travail.

Cas n° 2 : évaluation ex-post d'un programme du secteur des transports dans un pays de l'Asie du Sud

Les méthodes et démarches participatives suivantes seraient envisagées :

- Focus groupes et entretiens semi-structurés avec les principales parties prenantes et personnes affectées, par exemple : petits entrepreneurs du secteur formel et informel, transporteurs, conducteurs, membres d'un syndicat du port et gros employeurs.

¹ NdT : « Seeking surprise : rethinking monitoring for collective learning in rural resource management » de I. Guijt (2008), thèse de doctorat, Wageningen, Pays-Bas.

- Sur l'impact des routes secondaires en zone rurale, sélection d'un ensemble de conditions et de communautés avec identification informelle des gagnants et des perdants, puis entretiens semi-structurés ou focus groupes avec les personnes affectées, ce qui faciliterait le recensement d'autres gagnants et perdants, l'estimation des délais, l'établissement de graphiques de tendances et de changement, d'une matrice de scores avant-après, d'un diagramme de causalité des changements et d'un score de liens de causalité.
- Estimation chiffrée des gains (et des pertes) et, s'il y a lieu, « interrogation du diagramme » pour un éclairage approfondi.

La production d'estimations chiffrées va à l'encontre de l'idée que les démarches participatives ne produisent que des données qualitatives. On constate pourtant depuis dix ans une progression des données quantitatives générées par des démarches participatives, ce qu'on appelle en anglais les « *parti-numbers* », ou « chiffres participatifs » (voir Chambers, 2007).

Cas n° 3 : évaluation du soutien de donateurs à une commission anti-corruption dans un pays d'Afrique

Dans un domaine sensible comme la corruption, l'évaluation aurait plus qu'habituellement le caractère d'une intervention et serait traitée comme une opportunité. Cette réflexion resterait au cœur de l'évaluation afin d'accroître les bénéfices et de minimiser les dommages. Une première étape consisterait à recenser les formes et les niveaux (haut niveau, bas niveau) de corruption et les possibles liaisons causales entre le soutien apporté par les donateurs et les changements éventuellement constatés. Les ONG de défense des droits de l'Homme et les journalistes seraient des sources d'information privilégiées.

Avec les participants aux voyages d'études et les membres de la commission :

- Discussion réflexive en focus groupes (mais voir les entretiens individuels ci-dessous).

Pour une corruption de bas niveau, si des connexions plausibles étaient apparues, recherche de diverses sources de données et d'éclairage :

- Fiches d'évaluation complétées par les citoyens ou équivalent
- Focus groupes de personnes mécontentes
- Discussions informelles dans les échoppes à thé, etc.

Pour une corruption de haut niveau, la méthodologie accorderait un rôle privilégié :

- aux entretiens privés informels en tête-à-tête sans prise de notes. Pour les faciliter, dans la meilleure tradition de l'innocente mais machiavélique « *facipulation* » (facilitation à des fins manipulatrices), les évaluateurs disposeraient d'un budget à dépenser dans les bars.

Quelques remarques pour conclure

Une question a été posée sur la quantification du non-quantifiable. Les méthodes participatives ont la capacité, méconnue pour l'essentiel, de produire des chiffres qui peuvent être également commensurables et traités comme toute autre statistique. En formulant des appréciations, en procédant à des estimations et en exprimant certaines valeurs, les individus quantifient le qualitatif. Il est grand temps de reconnaître le

potentiel de ces méthodes. Comme toujours, il y a des questions éthiques. Bien déployées, les méthodes participatives peuvent être avantageuses pour tous – elles peuvent donner de l'autonomie aux populations et apporter des éclairages crédibles et fiables aux dirigeants politiques.

Les problématiques de pouvoirs et de l'économie politique des méthodologies ont été évoquées au cours de cette conférence. Quelqu'un a exprimé la crainte que les interventions soient choisies parce qu'elles se prêtent à des méthodes d'évaluation d'impact auxquelles on attribue une certaine rigueur. Je ne peux pas me prononcer sur ce point. Mais si, par exemple, cette approche produisait un biais en faveur de programmes dotés de moyens simples, standard et à action rapide aux dépens d'autres programmes plus complexes, à horizon plus long, pluralistes, participatifs et favorisant l'autonomie, le coût d'opportunité pour les pauvres pourrait être élevé.

Enfin, il est frappant de constater le peu de cas qui a été fait dans cette conférence de la primauté et des capacités des pauvres. L'objectif de l'évaluation d'impact est d'apprendre et d'apporter des changements qui leur rendent la vie plus facile. Nous avons besoin pour cela de méthodes mixtes et de pluralisme. L'évaluation d'impact peut et doit faire appel à une multiplicité d'outils et de démarches.

Quels qu'ils soient, ces outils et démarches doivent toujours reconnaître que ce sont les individus qui vivent dans la pauvreté, ceux qui sont vulnérables, marginalisés, qui sont les meilleurs juges et les autorités premières sur leur vie et leur moyens d'existence et sur la manière dont ils ont été affectés. Nous savons maintenant, ce que nous ignorions il y a 20 ans, que leurs capacités d'analyse sont très supérieures à ce que nous pensions. Nous savons qu'« ils sont capables ». Pour faciliter leur analyse et leur permettre ainsi de gagner en autonomie, nous disposons aujourd'hui de toute une panoplie de méthodologies participatives. Nous devons y recourir davantage et mieux et nous devons inlassablement réitérer cette injonction : demandez-leur !

Référence

Chambers, Robert (2007) « Who Counts? The Quiet Revolution of Participation and Numbers », *IDS Working Paper 296*, Institute of Development Studies.

Quelques réflexions sur les essais randomisés pour l'évaluation du développement : présentation à la conférence du Caire sur l'évaluation

*Dean Karlan**
Université de Yale
Innovations for Poverty Action
Jameel Poverty Action Lab

Résumé

Il nous a été demandé de présenter des approches méthodologiques précises pour l'évaluation de trois interventions fictives. Cet article saisit l'occasion de ce forum pour évoquer trois idées fausses sur les essais randomisés. Premièrement, personne n'affirme que les essais randomisés se prêtent à tous les cadres et à toutes les questions. Tout le monde est d'accord que la première des priorités consiste à poser la bonne question. Deuxièmement, le choix de ce que l'on mesure et comment, par des méthodes qualitatives ou participatives ou par des méthodes reposant sur des enquêtes quantitatives ou des données administratives, n'a aucun lien avec la décision de procéder ou non à un essai randomisé. Troisièmement, les essais randomisés se prêtent à l'évaluation de processus complexes et dynamiques, sans se confiner aux interventions simples et statiques. Les évaluateurs doivent s'efforcer de répondre aux questions les plus importantes pour les décisions futures et d'apporter les réponses les plus fiables possibles. La fiabilité est renforcée par des essais randomisés lorsqu'ils sont réalisables, en prêtant attention à la théorie sous-jacente et en réalisant des essais pour déterminer les raisons du succès ou de l'échec des interventions afin de pouvoir transférer au mieux les connaissances acquises à d'autres cadres.

1. Introduction

Pourquoi évaluer ? Trois raisons s'imposent naturellement : pour savoir où affecter des ressources limitées, pour savoir comment améliorer les programmes et pour inciter ceux qui ont des fonds à donner ou investir davantage.

Je voudrais commencer par un exercice mental inspiré du philosophe utilitariste Peter Singer. Porteriez-vous secours à un enfant en train de se noyer dans un lac s'il vous en coûtait 100 dollars parce que cela abîmerait vos vêtements ou vous ferait manquer un rendez-vous ? La plupart des gens répondent oui à cette question. Mais enverriez-vous immédiatement 100 dollars à une ONG dans un pays pauvre pour sauver un enfant ? De nombreuses personnes répondent non, au prétexte que personne ne sait vraiment si cette somme peut sauver un enfant ou si elle sera gaspillée. C'est une excuse courante à l'inaction. L'évaluation réfute cette excuse.

On a souvent opposé les essais contrôlés randomisés (ECR) à d'autres méthodes d'évaluation des programmes. Mais bien souvent, cette stricte ligne de partage entre les approches expérimentales et les autres est factice et masque l'association de méthodes expérimentales et qualitatives qui peut caractériser une bonne évaluation. Après avoir exposé quelques idées fausses couramment rencontrées sur la méthode de mesure, la capacité attributive et la faisabilité des ECR, je présenterai trois exemples de programmes de développement courants – allocation sociale conditionnelle, infrastructures et mesures anticorruption – et les circonstances dans lesquelles une stratégie d'évaluation doit ou non faire appel aux ECR.

* Contact : dean.karlan@yale.edu.

2. Quelques idées fausses sur les ECR

La première méprise est qu'il faut choisir entre une évaluation qualitative et un ECR, cette opinion reposant sur un spectre erroné de rigueur « attributive », qui place les ECR à une extrémité et les méthodes qualitatives à l'autre. En réalité, il n'y a pas de contradiction entre les méthodologies qualitatives et les ECR. En fait, une bonne évaluation au moyen d'un ECR passe souvent par une analyse approfondie du fonctionnement du programme, de son schéma initial, de sa théorie de changement, de la participation des bénéficiaires, etc.

Pour clarifier les débats sur les méthodes d'évaluation, il est impératif de distinguer nos échanges autour du recueil des données et de la mesure des résultats – ce qu'il faut mesurer, comment et qui inclure dans le processus – des moyens d'établissement des causalités entre les résultats et l'intervention. Les ECR établissent les relations de causalité en produisant une mesure du contrefactuel : ce qui se serait passé en l'absence du programme ou de la politique. Tout comme dans les essais médicaux, ils le font en assignant aléatoirement des individus au groupe traité et au groupe de contrôle, de façon qu'hormis le programme randomisé ou l'offre et sous réserve d'un échantillon de taille suffisante, les groupes soient identiques au plan des caractéristiques observables et non observables.

2.1. Établissement de la causalité

L'assignation aléatoire est utile en raison du biais de sélection, autrement dit parce que les participants au programme sont souvent différents des non-participants. Si on comparait les individus susceptibles de participer à un programme mais qui choisissent de ne pas le faire, on comparerait en fait deux groupes d'individus potentiellement très différents. On imagine aisément que ces groupes pourraient présenter des différences importantes, mais difficilement mesurables. Les participants au programme pourraient avoir davantage envie d'améliorer leur situation ou être plus autonomes ou plus instruits. Ils pourraient aussi avoir plus de temps libre. Les chercheurs s'efforcent souvent de neutraliser ces différences, mais il y a inévitablement des variables omises ou d'autres, comme la motivation, qui sont difficiles à mesurer. Du fait de ces différences, les estimations de l'impact de l'intervention sont biaisées, puisque les écarts de résultats entre le groupe traité et le groupe de contrôle peuvent résulter de ces caractéristiques non observées, et non de l'intervention.

2.2. Données et mesures

Les mesures quantitatives de résultats sont utiles aux évaluations parce qu'elles permettent aux chercheurs d'établir la signification statistique de l'impact des programmes. Mais les ECR n'imposent pas de méthode particulière de collecte des données. Ils peuvent utiliser des données quantitatives aussi bien que qualitatives, et ils les associent souvent au sein d'une même évaluation. Ils peuvent faire appel à des méthodes issues de l'économie, de la sociologie et de la psychologie ou d'autres disciplines et à des processus participatifs permettant l'expression des populations locales (voir par exemple Chattopadhyay et Duflo, 2004, qui ont constaté une plus grande propension des femmes au Bengal Occidental à participer au processus d'élaboration des politiques si le chef du conseil de leur village est une femme), entre autres, et même des « personnes extérieures » que Chambers évoque ici (Karlan et Zinman, 2009).

Une autre erreur est de penser que les adeptes des ECR affirment qu'ils peuvent et doivent être réalisés pour chaque programme. Les ECR sont un important outil de recherche parce que la causalité qu'ils établissent offre une mesure rigoureuse de l'impact du programme et aide ainsi à déterminer s'il y a lieu de le reproduire ailleurs et comment l'améliorer. Cependant, il n'est pas toujours possible de réaliser un ECR. L'opportunité d'un ECR dépend de la situation mais aussi de la question posée. Et

comme l'indique l'article de Ravallion (2009), il ne faut jamais commencer par la méthodologie pour se demander ensuite quelle question poser. Les évaluateurs doivent d'abord définir les questions auxquelles il faut apporter une réponse, puis réfléchir à l'outil le plus approprié pour cela. Lorsqu'ils sont réalisables, les ECR produisent l'estimation de l'impact du programme la moins biaisée, mais le fait qu'ils soient possibles ne signifie en aucun cas qu'il faut en conduire systématiquement. En l'absence de stratégie d'identification commode, les ECR sont assurément le moyen le plus pratique d'établir un cadre de recherche crédible.

2.3. *Approches créatives dans les ECR*

Tout en insistant sur le fait que les ECR ne peuvent fonctionner partout, j'ajoute qu'avec un peu de créativité, ils sont réalisables dans de nombreuses situations qui ne s'y prêtent pas a priori. Ainsi, les interventions peuvent souvent tirer profit des contraintes de mise en œuvre et randomiser au niveau de la communauté ou d'un autre échelon géographique au lieu de sélectionner aléatoirement des individus pour les assigner aux groupes de contrôle et de traitement. Plusieurs évaluations mesurant l'impact de la microfinance utilisent cette approche. Dans d'autres cas, un ECR peut exploiter les différences d'intensité de l'effort de promotion d'un programme auprès de différentes régions (visée d'encouragement). Pour un ECR, le critère décisif est la taille de l'échantillon, qui doit être fractionnable en suffisamment d'unités pour pouvoir mesurer les effets directs de diffusion et les effets d'équilibre général. Sous réserve d'une bonne préparation et à condition que les effets ne soient pas trop agrégés (par exemple au niveau d'un pays), des ECR bien structurés peuvent mesurer les impacts directs de l'intervention et les effets de diffusion positifs et négatifs sur d'autres groupes que les bénéficiaires directs. Ce sont en fait ces ECR qui produisent les comptes rendus les plus intéressants, parce qu'ils permettent de mieux comprendre comment une idée fonctionne sur un individu donné, mais aussi comment elle fonctionnera à plus grande échelle avec des effets directs et indirects.

2.4. *Intervention statique ou dynamique*

L'idée que l'intervention doit être homogène et statique est une autre méprise courante. De fait, les interventions « émergentes, complexes » ou « complexes », telles celles examinées ici par Rogers (2009), ne sont pas plus difficiles à gérer pour un ECR que pour une autre méthode. Les arguments qui suggèrent que la complexité et un processus dynamique causent des ravages avec un ECR passent à côté de la nature exacte du résultat d'un ECR. Un ECR aide simplement à générer un groupe de comparaison objectif auquel comparer les changements. L'intervention elle-même peut être statique et simple, ou complexe et dynamique. Dans ce dernier cas, l'évaluation est bien entendu décrite comme telle : on évalue un processus, une opportunité associée à certaines ressources, une intervention dynamique et fluide qui a été menée d'une certaine façon, etc. L'important c'est que l'évaluation porte sur le processus, et non sur les activités individuelles qui entrent dans le déploiement du programme. Si le projet devait fonctionner, ce qu'il faut reproduire c'est le processus de la mise en place des ressources, de la facilitation de leur emploi, etc. Le principe est très proche de nombreuses interventions de développement communautaire dans lesquelles des ressources comme la formation et une assistance technique individualisée sont apportées aux communautés et des exercices de facilitation sont mis en place pour les aider à se développer et à prospérer.

Nous conduisons justement une évaluation de ce type, recourant à un ECR avec outils qualitatifs et quantitatifs, du *Hunger Project* au Ghana, et d'un projet de développement communautaire participatif en Sierra Leone. Bien entendu, il faut bien comprendre que ce qui est évalué ici ce n'est pas une intervention précise mais un processus collaboratif. On ne peut pas savoir a priori quels moyens les acteurs en question sélectionneront, ni anticiper que le même processus donnerait lieu ailleurs à des choix identiques. Les enseignements d'une évaluation de ce type portent donc sur les changements que l'on peut attendre d'un tel processus – non pas des choix et des investissements précis que

font les acteurs, mais du processus de facilitation ou de financement des villages qui accompagne l'élaboration de leur propre programme. Cela dit, si les responsables du programme souhaitaient mesurer les impacts individuels des activités qui composent l'intervention, un ECR pourrait être structuré de façon à produire des résultats discrets à partir d'interventions complexes. Il faudrait pour cela opérer une variation aléatoire des composantes de l'intervention en groupes de traitement multiples. On comparerait probablement l'impact d'un ensemble de services, avec ou sans l'interaction d'une ou de plusieurs composantes additionnelles.

Les ECR offrent ici un important avantage par rapport à d'autres méthodes parce qu'ils peuvent surmonter les biais de sélection inhérents à de nombreux programmes sociaux et traiter les impacts directs des différentes activités entrant dans un protocole de traitement multiple. Par exemple, si lors de l'évaluation d'une formation à la gestion d'entreprise on constatait une augmentation des bénéficiaires, en particulier parmi ceux qui tenaient bien leur comptabilité, cela signifie-t-il qu'il faut promouvoir la formation comptable ? La tenue de la comptabilité est peut-être une composante importante de la formation mais il est possible aussi que les entrepreneurs les plus performants tiennent naturellement leur comptabilité. Les ECR peuvent démêler ces questions en assignant les participants à une formation avec ou sans module de comptabilité.

L'autre méprise fréquente est que les ECR ne mesurent les impacts d'une intervention que sur la moyenne de la population et ne tiennent pas compte des impacts différentiels sur ses différents segments. En fait, à condition d'avoir un échantillon de taille suffisante et un plan d'échantillonnage couvrant une diversité d'individus susceptibles d'accéder au programme, un ECR peut permettre d'identifier les groupes pour lesquels l'impact du programme est le plus important et ceux pour lesquels il est insignifiant, voire négatif. Ainsi, un ECR mesurant l'impact d'un programme de formation à la gestion de micro-entreprises au Pérou a montré de manière inattendue que les entreprises qui n'étaient pas intéressées par une formation complémentaire bénéficiaient en fait davantage du programme (Karlan et Valdivia, 2008).

3. Trois exemples

L'exposé de Ravallion donne une excellente vue d'ensemble des types de questions qu'il faut poser au début du processus d'évaluation pour définir l'objectif et le champ de l'évaluation, et donc les questions de recherche essentielles. Comme il l'indique, la possibilité d'une randomisation dépend de l'unité d'assignation. Ces trois exemples donnent un excellent éventail sur ce point précis. Je proposerai tout d'abord un protocole général d'évaluation de ces trois interventions, puis pour chacune d'elles, j'indiquerai comment répondre à des questions subsidiaires sur des points particuliers de mise en œuvre par des essais randomisés, même si l'évaluation de l'impact global emploie d'autres méthodes. Sans se substituer à l'évaluation globale non expérimentale, ces propositions peuvent offrir des méthodes intéressantes pour produire des données précises et objectives afin d'éclairer d'importantes questions de mise en œuvre pour l'avenir.

Concernant le premier exemple, l'allocation sociale conditionnelle, la méthode d'évaluation d'impact recommandée est l'essai contrôlé randomisé avec recueil de données quantitatives et qualitatives. Ce type d'essai a été réalisé dans plusieurs pays. Lorsque les États disposent de ressources limitées pour déployer les programmes d'allocations sociales conditionnelles (ASC) à grande échelle, la randomisation permet une distribution progressive des prestations particulièrement équitable et transparente. De récentes recherches ont montré que, bien structurés, les ASC peuvent être un bon moyen d'atteindre d'importants objectifs de politique publique. Cependant, la question du déploiement optimal de ces programmes est loin d'être réglée. On s'interroge par exemple sur la fréquence souhaitable des paiements, sur l'opportunité d'ajouter des services d'épargne et de faire coïncider le versement de la subvention avec les dépenses

d'éducation².

Pour le deuxième exemple, qui concerne les infrastructures, plusieurs protocoles sont techniquement possibles, mais ils requièrent un degré variable d'engagement de la part des fonctionnaires qui gèrent les programmes. J'aborderai la réhabilitation des ports, les routes principales et les routes secondaires. Malheureusement, on attend souvent qu'il soit trop tard pour demander une évaluation. Quelle que soit la méthode retenue, il est de loin préférable de préparer l'évaluation à l'avance, de délimiter soigneusement les objectifs et d'être exhaustif quant aux résultats qu'il convient de mesurer et aux méthodes de mesure.

L'évaluation de la réhabilitation d'infrastructures portuaires et d'un réseau de routes principales et secondaires pourrait privilégier les méthodologies d'évaluation de processus. La première étape consistera à établir un cadre logique, avec des objectifs chiffrés concernant par exemple le délai d'attente exprimé en jours et le nombre de jours jusqu'au transport, le coût du transport maritime et du transport terrestre, la valeur et la quantité des produits transportés et le nombre de bateaux, de camions et de voitures qui entrent et qui sortent. On peut faire appel à des outils économétriques, en fonction des effets différentiels sur les secteurs d'activité et des tarifs douaniers, par exemple. Il s'agit d'un simple suivi de programme et il est important à la fois pour gérer la mise en œuvre et rendre compte des résultats.

Dans ces situations, les ECR peuvent aider à répondre à des questions portant sur des aspects critiques de la théorie de changement du programme, mais ils ne porteront probablement pas sur l'intervention tout entière. Pour la réhabilitation d'installations portuaires dans un pays en développement, la question déterminante pourrait être la suivante : « une baisse des coûts de transport permettra-t-elle un gain de croissance de l'industrie en zone rurale ? » Dans ce cas, on pourrait envisager un ECR qui subventionne aléatoirement les coûts de transport dans certaines régions pour examiner le changement induit dans l'activité économique.

Il est techniquement possible de répondre à d'importantes questions de politique publique sur l'impact de programmes de routes secondaires au moyen d'un ECR, mais cela nécessitera probablement une volonté affirmée de la part des dirigeants politiques. Cet exemple renferme de considérables enseignements pour l'action publique d'autres pays, mais il nous paraît aussi politiquement difficile à accomplir. S'il y a assez de routes et si la topographie et les coûts de construction le permettent, on peut envisager de randomiser la construction progressive des routes. Imaginons un plan décennal d'amélioration ou de construction de routes secondaires rurales.

La randomisation de l'ordre des interventions est à la fois (a) équitable et (b) aisément évaluable. Elle pourrait être exécutée en incorporant la hiérarchie des routes au plan décennal si certaines routes sont plus importantes pour des raisons économiques et géographiques (ou politiques). Des dirigeants politiques entreprenants pourraient comprendre que dans ce contexte, un ECR évite tout favoritisme politique dans l'établissement du calendrier des interventions (c'est-à-dire que les routes seraient affectées à la tranche 1 ou à la tranche 2 dans un processus délibératif transparent, puis l'ordre de construction à l'intérieur de chaque tranche serait tiré au hasard, et donc juste). Dans ce cas, l'ECR est une facette du protocole d'évaluation. On pourrait aussi recourir à des méthodes économétriques, notamment aux approches de la double différence, à l'analyse avant-après ou à une comparaison transversale des routes construites et des routes non construites (en prenant par exemple des villes situées à 7 km de la route réparée ou construite ou 7 km d'une route non réparée ou non construite).

Le dernier exemple, celui des mesures anticorruption, ne se prête pas habituellement à

² Pour une bonne étude sur la conception de cadres décisionnels facilitant les choix éthiques, voir *Nudge*, par Richard Thaler et Cass Sunstein.

une évaluation de type attributif, même si le processus peut certainement produire des résultats mesurables comme des arrestations ou la révocation d'élus politiques. Mais on peut beaucoup apprendre sur la mécanique interne du comment et du pourquoi les fonctionnaires ont recours à des tactiques illicites. Et d'autre part, c'est un domaine où la transparence de la démarche d'évaluation est vraiment importante ! Un premier exemple est celui de la retransmission télévisée des audits des comptes des municipalités au Brésil. Un autre bon exemple est celui du travail d'Olken (2007) en Indonésie, qui nous renseigne sur l'effet comparatif de méthodes concurrentes de lutte contre la corruption, en associant une collecte de données qualitatives (corruption perçue et méthodes participatives de réunions de villages) et quantitatives (qualité des routes réelles).

4. Conclusion

Le dernier avantage d'un ECR est l'indépendance qu'il offre car il permet d'élaborer des tests statistiques rigoureux avant l'intervention, puis de laisser parler les données sur ce qui a marché ou non. En dernier ressort, l'objectif de l'évaluation doit être d'aider à décider de ce qu'il faut faire à l'avenir, pour les donateurs, qui doivent savoir où placer leur argent, pour les sceptiques, qui veulent constater que les programmes peuvent avoir un impact, et pour ceux qui les mettent en œuvre, qui ont besoin de savoir comment concevoir au mieux leurs programmes. Certains des travaux les plus intéressants font appel à des méthodes mixtes qui incorporent des méthodes qualitatives à des essais randomisés et utilisent des méthodes randomisées pour évaluer des processus dynamiques et complexes, tels que les programmes de développement communautaire.

Dans cet article, j'ai abordé deux questions fondamentales du débat sur les méthodes d'évaluation d'impact : la distinction entre l'objet de la mesure et l'objet de la comparaison. Lorsqu'on étudie ces questions, on voit que les ECR s'attachent à l'objet de la comparaison et qu'ils permettent d'intégrer de nombreuses méthodes participatives, qualitatives et quantitatives pour l'objet de la mesure. J'ai également tenté de dissiper quelques malentendus courants sur les extrêmes du débat. Les partisans des ECR eux-mêmes ne plaident pas pour qu'ils soient réalisés partout et pour chaque programme. Si je devais émettre une hypothèse, ce serait que les ECR drainent moins d'un pour cent des budgets d'évaluation. Je pense qu'ils devraient être réalisés plus souvent, mais pas à 100 % (ni à 99 %) non plus.

Références

Chattopadhyay, Raghavendra et Esther Duflo, 2004, « Women's leadership and policy decisions: evidence from a nationwide randomized experiment in India », *Econometrica*, vol. 72, n° 5, p. 1409–1443.

Karlan, Dean et Martin Valdivia, 2008. « Teaching Entrepreneurship: Impact of Business Training on Microfinance Institutions and Clients », Document de travail de l'Economic Growth Center de l'université de Yale.

Karlan, Dean et Jonathan Zinman, 2009, « Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts », *Review of Financial Studies*.

Olken, Benjamin, 2007, « Monitoring Corruption: Evidence from a Field Experiment in Indonesia », *Journal of Political Economy*, vol. 115, p. 200-249.

Ravallion, 2009, « Evaluating Three Stylized Interventions », *Journal of Development Effectiveness*, vol. 1, n° 3.

Rogers, Patricia J., 2009, « Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation », *Journal of Development Effectiveness*, vol. 1, n° 3.

Thaler, Richard et Cass Sunstein, 2008, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press, New Haven.

Évaluation de trois interventions stylisées

Martin Ravallion*

Groupe de la recherche sur le développement

Banque mondiale, Washington DC, 20433, États-Unis

Résumé

Comme les autres participants à une session de cette conférence, j'ai été invité à analyser le protocole d'évaluation de trois interventions stylisées : allocation sociale conditionnelle, programme dans le secteur des transports et commission de lutte contre la corruption. Cet article reprend mon intervention en développant certains points et en y ajoutant des références bibliographiques. Je commence par formuler quelques suggestions d'ordre général sur les aspects à envisager préalablement à toute évaluation avant d'illustrer ces remarques dans le contexte des trois interventions mentionnées.

1. Introduction

Les participants à cette session ont été invités à exposer les méthodes d'évaluation de trois interventions :

1. « Une allocation sociale conditionnelle dans un pays d'Amérique centrale, versée aux ménages dont les filles d'âge scolaire restent scolarisées et respectent certaines conditions d'assiduité et de résultats. »

2. « Un programme du secteur des transports dans un pays d'Asie du Sud comprenant la réhabilitation d'infrastructures portuaires, la remise en état de routes principales et de nouveaux investissements dans des routes secondaires en zone rurale. »

3. « Une commission anticorruption (CAC) dans un pays d'Afrique. Le programme prévoit une aide à l'élaboration de directives, la mise à niveau des infrastructures et des voyages d'étude. Des programmes comparables sont mis en œuvre dans six pays. »

Ces trois interventions présentent de nombreuses différences, mais la plus importante dans le contexte de l'évaluation réside dans la mesure dans laquelle elles peuvent être traitées comme des programmes ciblés, ce qui signifie que certaines unités observées (ménages, entreprises, villages, régions) reçoivent le programme, mais d'autres non. Une autre différence, liée à la précédente, est la mesure des effets secondaires, c'est-à-dire les effets (positifs ou négatifs) du programme sur des unités non ciblées. Le fait qu'un programme soit ciblé ne signifie pas évidemment qu'il n'a pas d'incidence sur des unités non ciblées. Les effets de diffusion peuvent constituer une importante source de biais dans les méthodes d'évaluation classiques³.

Aux deux extrêmes, une allocation sociale conditionnelle (ASC) est un programme ciblé, visant des ménages précis avec des effets de diffusion probablement modestes, alors qu'une CAC est généralement à l'échelle de l'économie, bien qu'elle puisse avoir des aspects ciblés, notamment lorsqu'elle « vise » des services particuliers de

* Les vues exprimées ici sont celles de l'auteur et ne doivent pas être attribuées à la Banque mondiale ou à une autre organisation affiliée. Contact : mravallion@worldbank.org. L'auteur remercie les participants à la conférence pour leurs remarques et Phil Keefer, Norbert Schady, Dominique van de Walle et Howard White pour les échanges intéressants et leurs observations judicieuses.

³ Voir par exemple Chen *et al.* (2009) dans le contexte des réponses des dépenses publiques locales aux programmes de développement des régions pauvres en Chine et Oduor *et al.* (2009) sur les effets secondaires de la promotion sociale des traitements de la malaria au Kenya.

l'administration ou des entreprises précises. Entre ces deux extrêmes, le programme du secteur des transports est un hybride ; la plupart de ses composantes sont ciblées, bien qu'il puisse avoir d'importants effets de diffusion.

Les programmes ciblés et les programmes non ciblés requièrent des outils d'évaluation différents même si les principes d'évaluation sont les mêmes – notamment la nécessité d'évaluer l'impact par rapport à un contrefactuel explicite. Un programme ciblé facilite les comparaisons d'observations dans lesquelles un échantillon sélectionné d'unités non ciblées est utilisé pour tenter d'inférer le contrefactuel dans certaines hypothèses identificatrices⁴.

Pour chaque cas, je commencerai par quelques questions clés sur le contexte de l'intervention avant d'aborder les questions d'évaluation. Naturellement, il est difficile d'être très précis sans avoir plus d'éléments sur le contexte.

Tout d'abord, quelques remarques d'ordre général intéressant les trois interventions.

2. Questions génériques

Les questions essentielles que j'aime poser en préalable à toute évaluation sont les suivantes :

- *Pourquoi cette intervention ?* Il est important de comprendre la motivation de l'intervention spécifique pour concevoir une évaluation utile, mais c'est aussi important plus généralement pour l'élaboration de bonnes politiques (ce qui est en fin de compte l'objectif ultime de l'évaluation). L'examen approfondi de la motivation de l'intervention peut même aboutir à une intervention différente.
- *Que savons-nous aujourd'hui de ce type d'intervention et quelles sont les lacunes les plus importantes dans nos connaissances ?* Il y a presque toujours une expérience passée qui peut être utile. Dès le départ, une étude attentive des données passées peut être révélatrice et influencer le protocole du programme et sa mise en œuvre, ainsi que les questions sur lesquelles portera l'évaluation.
- *Quel est le contrefactuel pertinent ?* Le contrefactuel classique est l'absence de programme, mais ce n'est pas toujours le plus intéressant pour les décideurs politiques qui affecteront souvent les mêmes ressources à un autre programme. Un programme peut sembler produire de bons résultats par rapport à l'hypothèse d'une absence totale de résultats, tout en donnant des résultats médiocres par rapport à une alternative faisable. Au plan formel, le problème de l'évaluation n'est pas fondamentalement différent si un autre programme est le contrefactuel ; en principe, on peut reproduire l'analyse relative au contrefactuel de l'absence d'intervention pour chaque alternative possible et les comparer. Cette procédure est assez rare dans le domaine de l'évaluation des projets de développement, mais elle l'est moins dans le secteur des soins de santé et des essais médicaux (où l'intervention existante est affectée au groupe de contrôle et l'intervention nouvelle au groupe de traitement).
- *Quels sont les résultats souhaités ? Dans quels délais ? Et quels sont les effets indésirables potentiels ?* Il est très important de connaître les objectifs et de savoir comment ils peuvent être traduits en résultats mesurables précis (quantitatifs ou qualitatifs). Il est tout aussi important de savoir dans quel délai les résultats (positifs et négatifs) sont attendus⁵.
- *Quels paramètres faut-il estimer ?* Les évaluations classiques ne s'attachent qu'à deux paramètres, l'impact moyen d'une intervention sur les unités qui ont la possibilité

⁴ Pour un complément d'analyse sur les hypothèses et méthodes utilisées pour les programmes ciblés et des références à la littérature (abondante), voir Ravallion (2008).

⁵ Voir King et Behrman (2009) pour une intéressante discussion sur ce point.

de la suivre (paramètre de l'« intention de traiter ») et l'impact moyen sur celles qui la reçoivent. Mais en général, les dirigeants politiques ne s'intéressent pas qu'à ces deux paramètres. Les questions suivantes sont également utiles : l'intervention fonctionne-t-elle comme prévu ? Quelles catégories d'individus en bénéficient et quelles catégories d'individus y perdent ? Quelle est la proportion de participants qui en bénéficie ? Que se passe-t-il quand le programme est déployé à plus grande échelle ? Comment le structurer autrement pour en renforcer l'impact ?

- *Quels sont les mécanismes de transmission attendus ?* L'évaluation d'impact classique peut être une « boîte noire » qui nous donne très peu d'informations sur les mécanismes d'impact ou d'absence d'impact d'un programme. Pour concevoir une évaluation susceptible d'éclairer cette boîte noire, il faut comprendre la motivation théorique d'un programme – les modes d'action précis par lesquels l'intervention est censée améliorer la vie des individus. (Je reviens plus loin sur cette évaluation « basée sur la théorie »). Les décideurs politiques ont souvent une conception assez vague de ces mécanismes et l'examen réalisé par des évaluateurs peut être utile en exposant la motivation de l'intervention et parfois en révélant qu'elle est douteuse.

- *Quelles sont les options méthodologiques dans ce cadre ?* L'évaluation s'assortit de contraintes (techniques, économiques, politiques et éthiques) qu'il faut recenser dès le départ. Des méthodes d'évaluation par ailleurs souhaitables peuvent être irréalisables dans un contexte particulier. Parmi les options possibles, le choix doit dépendre des réponses aux questions précédentes et non des préférences méthodologiques de l'évaluateur. Même si cela peut paraître évident, ce n'est pas en réalité la pratique courante ; trop souvent, l'évaluateur apporte avec lui sa panoplie d'outils favoris et choisit les questions auxquelles ces outils précis peuvent apporter une réponse (et non l'inverse). Parfois, les outils possibles sont remarquablement peu nombreux et l'évaluateur commence avec une seule méthode privilégiée (comme la randomisation, une méthode économétrique non expérimentale ou un outil qualitatif favori) et recherche les questions qu'il est possible d'aborder avec cette méthode. En règle générale, ce type de démarche ne produit pas les meilleures évaluations.

Les réponses à ces questions sont souvent essentielles pour le protocole d'évaluation, mais elles sont propres à chaque programme et à son cadre. Il n'existe pas de démarche appropriée pour toutes les interventions indépendamment du cadre. Cela posé, j'expose ci-après quelques réflexions sur la problématique d'évaluation propre à ces trois interventions relativement différentes.

3. Programme d'allocation sociale conditionnelle

C'est l'exemple le plus facile des trois, pour deux raisons. Tout d'abord, une ASC est un programme ciblé et on peut raisonnablement supposer que ses effets de diffusion sont très faibles. Ensuite (et en partie pour la première raison), de nombreuses recherches ont été conduites dans le domaine de l'évaluation des ASC, de sorte que nous disposons de connaissances assez étendues sur les problèmes posés par ces programmes et que nous savons comment les évaluer au mieux. Des lacunes demeurent cependant.

L'idée de base d'une ASC est que la famille bénéficiaire doit apporter la preuve d'une assiduité scolaire suffisante et (dans certains exemples) de soins de santé appropriés ; l'allocation n'est versée que si ces conditions (parfois appelées « coresponsabilités ») sont vérifiées⁶. Parmi les premiers exemples influents, citons le programme *Food-for-Education* au Bangladesh, le programme mexicain *PROGRESA (Programa de Educacion, Salud y Alimentacion)* (aujourd'hui appelé *Oportunidades*) et *Bolsa Escola* au Brésil. Les évaluations d'impact montrent que ces programmes apportent des bénéfices non

⁶ Le terme « conditionnelle » est impropre car la quasi-totalité des programmes de protection sociale et de transferts sociaux imposent des conditions aux bénéficiaires, mais le terme est uniquement appliqué à ce type de programme.

négligeables aux ménages pauvres, en termes de revenu actuel et de revenu futur, par des investissements plus élevés dans l'éducation des enfants et les soins de santé ; pour une récente revue des données issues des évaluations réalisées, voir Fiszbein et Schady (2009).

La motivation d'une ASC n'est pas aussi évidente qu'on pourrait le penser. Une question déjà ancienne est celle de la justification de la conditionnalité. Les insuffisances du marché du crédit, qui privent les ménages pauvres de la possibilité d'emprunter pour financer les études de leurs enfants, sont souvent citées comme une motivation d'une ASC, mais il peut y avoir de meilleures politiques pour résoudre ce problème, notamment des transferts sociaux non conditionnels (Das *et al.* 2005). Les questions relatives aux inégalités au sein des familles et à l'économie politique dominant souvent les autres motivations, y compris les insuffisances du marché du crédit.

L'efficacité de ces programmes dans des pays à faible revenu (la plupart des programmes et évaluations viennent de pays à revenu intermédiaire, en particulier d'Amérique latine) est un des aspects sur lesquels les connaissances sont insuffisantes. Ce déficit de connaissances est en passe d'être comblé grâce aux ASC en cours d'évaluation (en particulier en Afrique), et il est probable que de nombreux résultats seront produits dans les années à venir.

Nous avons besoin d'en savoir plus sur la composition du dispositif d'ASC pour déterminer comment fixer la composition précise des transferts (montant et nature des allocations sociales, en espèces ou en nature) et les incitations à un changement de comportement (quels comportements faut-il encourager). Des données montrent que dans le cas de *PROGRESA*, un report du primaire au secondaire, neutre sur le budget, de l'allocation pour scolarisation aurait relevé le niveau d'instruction en augmentant la proportion d'enfants qui poursuivent dans le secondaire (de Janvry et Sadoulet, 2006). Cela illustre ma remarque précédente : les évaluateurs doivent comprendre plus précisément les mécanismes d'impact des programmes pour que nous puissions donner des conseils plus avisés sur les moyens de les améliorer.

Des recherches intéressantes ont été conduites sur les effets comportementaux, mais elles se sont surtout attachées aux comportements voulus par les concepteurs des programmes (à savoir le respect des coresponsabilités). Nous avons besoin d'en savoir plus sur d'autres effets, y compris sur l'offre de main-d'œuvre et les décisions d'épargne des parents.

Une autre lacune du corpus des connaissances constitué par les recherches concerne les impacts sur le bien-être des changements comportementaux induits. Les résultats doivent couvrir la pauvreté actuelle et future. La mesure de la pauvreté actuelle pose moins de difficultés (sans pour autant être facile). Il peut être utile de s'intéresser aux « résultats intermédiaires » comme le travail des enfants et la fréquentation scolaire, mais nous en savons moins sur les résultats finaux. Certes, davantage d'enfants de familles pauvres fréquentent l'école élémentaire (par exemple) lorsque les conditions abaissent le prix de la scolarité. Mais apprennent-ils quelque chose d'utile ? Ou est-il trop tard pour que le programme ait un effet sur leurs capacités d'apprentissage ? Et qu'en est-il des impacts sur les revenus du travail post-scolaire ?

Les facteurs contextuels relatifs à l'offre revêtent une importance de premier plan dans les ASC. Les résultats décevants en termes d'apprentissage des enfants et de nutrition révélés par certaines évaluations peuvent être rapprochés de facteurs relevant de l'offre — mauvaise qualité des écoles et des établissements de soins — qui varient d'un lieu à l'autre. En l'absence d'amélioration de l'offre, les changements de comportement induits par une ASC peuvent être insuffisants pour obtenir des résultats sur le bien-être (Fiszbein et Schady, 2009).

Par ailleurs, nos connaissances présentent d'importantes lacunes quant aux impacts à plus long terme des ASC. Cela prendra naturellement du temps, mais c'est maintenant

qu'il faut investir dans les données. Hélas, les externalités en jeu et les difficultés consécutives de financement et de poursuite d'études des impacts à long terme sont loin d'être négligeables.

L'autre domaine dans lequel nos connaissances sont insuffisantes concerne la flexibilité des ASC du point de vue de l'ajustement des conditions d'ouverture des droits à l'évolution des besoins. C'est important si l'on veut que ces programmes servent de filet de sécurité, comme dans le contexte de la crise financière mondiale qui sévit aujourd'hui. On peut supposer qu'il est plus facile en cas de crise d'augmenter temporairement les allocations versées aux bénéficiaires actuels que d'accroître temporairement le nombre de bénéficiaires. Mais d'autres recherches sont nécessaires sur ce point et sur les moyens d'assouplir les programmes d'ASC.

Lorsqu'on envisage les options d'évaluation, il est important d'examiner les projets de déploiement échelonné du programme d'ASC. Même s'il n'est pas prévu de phase pilote, il ne sera pas toujours possible (pour des raisons techniques et budgétaires) de déployer le programme en une seule fois sur tout le territoire national et certaines régions pourront disposer du programme avant d'autres. Il est essentiel de comprendre la sélection des régions ciblées. Il faut rechercher les possibilités de créer un groupe de contrôle dans des régions non participantes comparables au plan de l'observation. Ou si une formule est appliquée pour sélectionner les régions prioritaires, on peut toujours envisager un protocole discontinu, qui détermine l'impact à proximité du plafond d'admission. L'expansion du programme permettra d'effectuer des comparaisons interrégionales, mais il pourrait être possible de déterminer les impacts plus tôt. Sur ce point, il peut être important de savoir si les régions de comparaison savent qu'elles rejoindront le programme et à quel moment, car cela peut produire des effets de contamination, les unités de comparaison étant affectées par les anticipations de participation au programme.

À bien des égards, l'évaluation initiale de *PROGRESA* est un modèle pour d'autres ASC. La caractéristique essentielle du protocole d'évaluation est qu'un groupe « exclu par randomisation » a servi de contrôle dans la période de déploiement progressif du programme. La longévité de ce programme (qui a survécu aux changements de gouvernement) et son influence dans la communauté du développement découlent en partie de l'effort important, et public, qui a été consenti pour l'évaluer. Un tiers des communautés de l'échantillon jugé admissible au programme a été choisi de manière aléatoire pour former un groupe de contrôle qui, dans une phase initiale, n'a pas bénéficié du programme tandis que les deux autres tiers en bénéficiaient. L'accès du public aux données produites par l'évaluation a permis plusieurs études intéressantes. Pour un panorama complet de la conception, de la mise en œuvre et des résultats de l'évaluation de *PROGRESA*, voir Skoufias (2005).

La conception de *PROGRESA* peut être améliorée. Comme nous l'avons dit plus haut, un inventaire plus complet des variables de résultats serait souhaitable. Des préoccupations demeurent quant à une possible contamination du groupe de contrôle, en particulier du fait d'effets d'anticipation (surtout lorsque, comme dans *PROGRESA*, le programme est déployé à l'échelle nationale au cours de la période d'évaluation et fait naturellement grand bruit). Le protocole d'évaluation aurait également pu faire mieux pour éclairer les questions relatives à la conception du programme, par exemple si l'allocation devrait être versée au niveau du primaire ou du secondaire. Les recherches menées par la suite ont également révélé l'importance des facteurs de l'offre, et une meilleure utilisation d'enquêtes sur les établissements scolaires et de soins aurait pu être utile.

4. Programme du secteur des transports

L'évaluation de programmes dans le secteur des transports doit envisager un certain nombre de questions contextuelles importantes. Logiquement, la première étape est de comprendre la nécessité de l'intervention, qui sera liée (entre autres) à l'histoire économique, à la géographie et à l'économie politique. A-t-on volontairement négligé

certaines ports et liaisons de transport, d'où le besoin de réhabilitation, et si oui, pourquoi? Les questions relatives aux finances publiques et aux effets de diffusion seront sans doute importantes. Comment le programme sera-t-il financé, et peut-on anticiper des effets de fongibilité ou d'adhérence (« *flypaper effect* », c'est-à-dire que malgré une fongibilité partielle, l'aide « colle » au secteur auquel elle est destinée) ? Évaluons-nous la bonne intervention ?

On peut anticiper des impacts à court et long terme. Il y aura des effets temporaires sur les revenus du travail (tant pour les travailleurs employés sur les projets que pour ceux qui pourraient bénéficier d'une tension des marchés du travail), ce qui peut être très important dans certains contextes, par exemple lorsque le projet de transport fait partie d'un plan de lutte contre la crise. La phase de construction s'accompagnera aussi de divers coûts, outre les coûts de construction, tels que les impacts sociaux et environnementaux (y compris des familles déplacées).

Cela étant, la majeure partie des impacts interviendra après la construction et le travail d'évaluation doit en tenir compte. Ces impacts peuvent être très divers et difficiles à quantifier. Les économies en coûts de transport (montants décaissés et valeur du gain de temps de transport pour les personnes et pour les marchandises) seront certainement de la plus haute importance. Pour les grands projets, il faudra probablement compter avec des effets (d'équilibre général) plus vastes sur l'activité économique, y compris sur sa géographie (sur laquelle je reviendrai).

L'évaluation devra considérer les implications pour les marchés et leur fonctionnement. Nous voudrions savoir si les marchés existants de marchandises échangées localement fonctionnent bien, ce qui permet de déterminer les impacts par le biais des prix. (L'évolution des prix est une des principales incidences sur le bien-être d'un projet réalisé dans le secteur des transports). Nous devons aussi déterminer si les marchés terrestres existants fonctionnent bien dans le cadre précis, ce qui peut permettre de déterminer les impacts à travers la capitalisation foncière. (Le modèle classique de Von Thunen prédit que les rentes foncières décroissent avec la distance au centre-ville, ce qui reflète les coûts de transport ; pour un exemple montrant comment utiliser ce type de modèle pour évaluer des projets du secteur des transports, voir Jacoby, 2000.)

Nous voudrions sans doute également savoir comment l'intervention affecte la géographie de l'activité économique, notamment le développement local du marché et des institutions. L'amélioration des routes en zone rurale, par exemple, attire-t-elle des marchés, des institutions et une nouvelle activité économique vers les régions pauvres en retard, ou encourage-t-elle une nouvelle concentration géographique de ces activités ? Comment les grands projets du secteur des transports modifient-ils la géographie de l'activité économique ?

Compte tenu de l'importance manifeste des coûts de transport dans les décisions d'implantation, on peut penser qu'un grand projet du secteur des transports modifiera le paysage économique, en bénéficiant à certaines régions et activités mais en conduisant peut-être d'autres au déclin. La Nouvelle Économie géographique prédit que la baisse des coûts de transport des produits (agricoles et non agricoles) permise par des investissements d'infrastructure accroît la concentration géographique d'autres activités (non agricoles) en zone urbaine pour exploiter les économies d'agglomération (Fujita *et al.* 2001, chapitre 7). L'évaluation d'un projet de routes rurales soutenu par la Banque mondiale au Vietnam a constaté que le projet stimulait le développement du marché local dans les régions pauvres — en déconcentrant partiellement l'activité économique ; voir Mu et van de Walle (2008). Il est bien possible que les impacts locaux en zone rurale sur la concentration géographique de l'activité diffèrent des impacts régionaux, y compris dans les zones urbaines.

Une question qui ne manquera pas de se poser est celle de savoir si l'évaluation doit exclusivement porter sur les résultats du point de vue de l'efficacité économique globale. Certains avancent parfois qu'une politique de développement ne doit pas se préoccuper

de l'équité de l'allocation spatiale de l'activité ni faire obstacle à l'allocation déterminée par le marché, et qu'elle doit utiliser d'autres instruments (« spatialement neutres ») pour résoudre les problèmes d'équité, y compris ceux qui ont une dimension géographique tels que le retard de développement des régions pauvres ; voir Banque mondiale (2009) pour une argumentation de ce type. Si l'on appliquait ce point de vue aux programmes du secteur des transports, on ne se préoccuperait que de leur efficacité, mesurée par leur impact sur le revenu moyen.

Je doute cependant que de nombreux dirigeants politiques accepteraient ce point de vue et ils auraient raison. Ils savent qu'ils n'ont pas tous les instruments de politique nécessaires pour un traitement spatialement neutre des considérations d'équité⁷. De fait, dans les pays pauvres, la localisation est une des dimensions les plus utilisées pour recenser les pauvres à des fins de redistribution, faute de meilleures informations sur les niveaux individuels de bien-être. On peut penser que pour un certain temps encore, dans les pays en développement, les évaluations de projets de transports à des fins de politique publique devront considérer les aspects d'équité, et dans les dimensions non relatives aux revenus également (qui comprennent par exemple les impacts sur la santé infantile résultant du meilleur accès des mères aux soins de santé permis par l'amélioration des transports).

Le protocole d'évaluation d'un programme de transport doit distinguer les composantes qui peuvent être traitées comme des programmes ciblés (tels que les routes locales en zone rurale) de celles qui sont susceptibles d'avoir des impacts beaucoup plus spatialement dispersés (routes principales par exemple) et requièrent donc une approche à l'échelle d'un secteur ou de l'économie.

L'évaluation doit être conçue différemment pour les composantes ciblées et non ciblées. Pour les premières, telles que les améliorations du réseau routier en zone rurale, nous voudrions déterminer la « zone d'attraction » qui recouvre les impacts probables (van de Walle, 2009). Nous avons également besoin de comprendre comment le placement géographique a été déterminé et de trouver des zones d'observations similaires qui ne reçoivent pas de nouvelles routes rurales ou de dépenses de réhabilitation. Il est presque certain que des observations longitudinales seront nécessaires (pour permettre un estimateur « de matching en double différence » ; voir Ravallion, 2008, pour un complément d'analyse et des exemples). Les évaluateurs devront également déterminer les sources probables de biais de sélection. Si la sélection repose essentiellement sur des observables, il doit être possible de recueillir les données permettant de corriger ce biais de manière satisfaisante. Si la sélection est opérée sur des critères non observables, il faudra envisager des méthodes statistiques telles que l'estimation des variables instrumentales (une randomisation est rarement possible pour des raisons évidentes).

Si le projet de transport est important, sa zone d'attraction le sera aussi et il pourrait être difficile de trouver de bonnes régions de comparaison, non touchées par le programme. Nous devons nous tourner vers des outils assez différents⁸. À un extrême (en termes d'agrégation), des régressions de croissance internationales ont été utilisées pour déterminer les impacts des infrastructures, notamment de transport ; voir par exemple Calderon et Servén (2008). Des données géographiques infranationales liées à des données sur les ménages et les entreprises peuvent permettre une analyse plus fine ; voir par exemple Jalan et Ravallion (2002), qui constatent que les routes rurales en Chine ont un impact positif sur le processus de micro-croissance. Les modèles d'équilibre général calculables spatiaux, issus de la Nouvelle Économie géographique (NEG), pourraient bien à l'avenir jouer un rôle important (Fujita *et al.* 2001).

⁷ Il est depuis longtemps reconnu en économie que des conditions très fortes doivent être vérifiées pour une stricte séparation des instruments d'équité et d'efficacité dans la réalisation d'objectifs généraux de bien-être social. Il est généralement admis que ces conditions (en particulier la faisabilité de transferts monétaires non facteurs de distorsion) ne se vérifient pas en pratique.

⁸ On trouvera une intéressante vue d'ensemble des options méthodologiques dans Oosterhaven et Knaap (2003).

Ces modèles sont conceptuellement bien adaptés à l'évaluation des impacts de grands projets de transport, même si ce sont aussi des modèles complexes, exigeants en termes de données et de calibration. L'Union européenne a développé un modèle EGC spatial pour évaluer l'amélioration des transports ; pour une application dans le contexte de l'évaluation d'un projet de grande ampleur dans les transports (un projet ferroviaire aux Pays-Bas), voir Knaap et Oosterhaven (2000).

Il faut reconnaître que les anciens modèles de la NEG ont émis des hypothèses assez irréalistes dans la modélisation des coûts de transport, et il est possible qu'elles ne soient pas étrangères aux enseignements tirés pour les impacts d'améliorations dans le secteur des transports⁹. On peut penser que des modèles opérationnels basés sur des hypothèses plus réalistes seront construits à l'avenir.

5. Commission anticorruption

Il s'agit d'un organe indépendant investi de pouvoirs exceptionnels pour enquêter sur la corruption et engager des poursuites, qui est généralement rattaché au chef de l'État ou au parlement¹⁰. Deux grandes questions contextuelles devront être considérées dans l'évaluation d'une CAC. Premièrement, il faut déterminer si l'intervention traite les causes connues de la corruption. Les causes profondes de la corruption résident-elles dans les mécanismes d'application de la loi ou dans les incitations des acteurs politiques à surveiller les fonctionnaires ? Doit-on plutôt parler de réformes des institutions fiduciaires, surtout de la gestion du secteur public ? Deuxièmement, les évaluateurs doivent savoir précisément ce qui motive cette intervention. Est-ce simplement que les donateurs ont envie de voir sanctionner les fonctionnaires corrompus ou qu'un leader politique souhaite réprimer l'opposition, ou bien l'intervention résulte-t-elle d'une analyse plus approfondie du problème ?

Il est possible de concevoir des évaluations pour déterminer l'impact sur la corruption des types d'outils utilisés par une CAC. L'étude d'Olken (2007) sur la corruption dans le cadre des projets routiers en Indonésie en donne un exemple innovant, qui recourt à une approche aléatoire. Olken a constaté qu'une augmentation des contrôles réalisés par le gouvernement central réduit l'ampleur des « dépenses manquantes » mesurées par la différence entre les coûts du projet officiellement enregistrés et les estimations indépendantes réalisées par des ingénieurs. (À l'inverse, Olken (2007) a observé que les approches participatives du suivi avaient peu d'impact.)

Des études évaluatives de ce type pourraient avoir un apport intéressant, mais elles ne déterminent pas l'impact d'une CAC en tant que telle. L'étude d'Olken montre que le suivi par des auditeurs publics peut aider à lutter contre la corruption en Indonésie ; cependant, selon au moins une évaluation (peut-être obsolète), les efforts de type CAC consentis par l'Indonésie ne semblent pas avoir réussi à faciliter ce suivi et ces sanctions au-delà de quelques arrestations, largement médiatisées, d'individus notoirement corrompus associés au régime politique précédemment au pouvoir (Sherlock, 2002).

La CAC type n'est pas un programme ciblé comme les ASC ou certaines composantes d'un programme de transport évoqué plus haut. Une expansion progressive de la couverture géographique ou ministérielle peut être possible en attribuant les pouvoirs de la CAC à certaines collectivités locales ou à certains ministères plutôt qu'à d'autres. Si l'expansion progressive est possible et si les effets d'anticipation ne risquent pas d'être trop marqués, la comparaison des résultats de tests de corruption objectifs (tels ceux utilisés par Olken, 2007) entre des administrations observables similaires (dans des

⁹ Je fais ici référence à la « fonction des coûts de transport de type iceberg », qui implique que les prix des marchandises livrées augmentent de façon exponentielle avec la distance parcourue, ce que contredisent les données ; pour un complément d'analyse, voir McCann (2005).

¹⁰ Pour un récapitulatif de l'histoire et des performances des CAC, voir Rose-Ackerman (1999, chapitre 9).

zones de collectivités locales différentes), qui se différencient par le fait que certaines sont sous la surveillance de la CAC et d'autres non, pourrait aider à déterminer l'impact de la CAC. Cependant, même en faisant abstraction des questions de faisabilité (le gouvernement pourrait préférer donner carte blanche à la CAC dès le départ), les effets d'anticipation et les biais de sélection posent de sérieux problèmes. (Le signe du biais de sélection est ambigu en l'absence d'autres informations sur le cadre ; un président réellement désireux de lutter contre la corruption choisirait sans doute les lieux ou les secteurs où le problème est jugé le plus grave ; un président qui se servirait de la CAC à des fins politiques effectuerait des choix différents et pourrait même éviter les foyers de corruption.)

Une approche plus prometteuse (peut-être complémentaire) serait de considérer les voies d'impact possibles de la CAC, en partant d'un modèle théorique plus ou moins explicite des raisons de la corruption. L'évaluation pourrait alors porter sur ce qui semble constituer les maillons clés de la chaîne causale attendue, comme le type d'« évaluation basée sur la théorie » analysée dans Weiss (2001), Rogers (2009) et White (2009). Klitgaard (1988), par exemple, a avancé que la corruption résulte de trois facteurs : un monopole sur une ressource, un contrôle discrétionnaire des fonctionnaires sur l'accès à cette ressource et l'absence ou l'échec de mécanismes pour responsabiliser ces fonctionnaires. On pourrait alors commencer par déterminer comment la CAC aborde chacune de ces composantes en modifiant les incitations des fonctionnaires dans l'environnement institutionnel concerné. Il faudra pour cela évaluer les incitations offertes individuellement aux fonctionnaires et déterminer en quoi elles ont été modifiées par la CAC. Ce n'est pas une tâche facile, compte tenu du peu d'incitations à dire la vérité à un évaluateur. Mais au moins, en basant l'entreprise d'évaluation sur la compréhension des causes de la corruption (qui sont bien plus profondes que l'« immoralité » des fonctionnaires), il serait peut-être possible de déterminer si la CAC a un réel impact sur le problème.

Les CAC aiment mesurer leur efficacité à l'aune du nombre de fonctionnaires poursuivis. Cette mesure peut être trompeuse car la CAC peut impliquer un ciblage sélectif des opposants politiques. Compte tenu de cette possibilité, l'évaluation pourrait utilement se concentrer directement sur le soutien ou l'opposition politique à l'égard de la CAC. Des données pourraient être recueillies sur l'affiliation politique des cibles de la CAC, associées à des enquêtes auprès du personnel sur les contacts avec les hommes politiques, et les motifs de départ des personnels n'exerçant plus leurs fonctions. Il faudra ensuite effectuer des enquêtes de suivi (après l'enquête de référence), qui sonderont les ménages, les entreprises et les fonctionnaires sur le degré de politisation de la CAC par rapport à d'autres institutions comme la Banque centrale, le ministère des Finances ou le Parquet.

6. Conclusion

L'art de bien évaluer consiste à poser les bonnes questions d'emblée, motivées par les déficits de connaissances, et à adapter les données et l'analyse pour répondre à ces questions dans le contexte considéré. Il est impossible, dans un article comme celui-ci, d'anticiper toutes les questions les plus importantes pour une évaluation donnée et toutes les méthodes d'évaluation possibles ; les unes et les autres dépendront du contexte précis. Nous espérons cependant que cet article aura au moins apporté quelques bases de réflexion.

Références

Banque mondiale, 2009, *Reshaping Economic Geography*, Banque mondiale, Washington DC.

Calderon, Cesar et Luis Servén, 2008, « Infrastructure and Economic Development in Sub-Saharan Africa », Policy Research Paper 4712, Banque mondiale, Washington DC.

- Chen, Shaohua, Ren Mu et Martin Ravallion, 2009, « Are There Lasting Impacts of Aid to Poor Areas? Evidence from Rural China », *Journal of Public Economics*, vol. 93, p. 512-528.
- Das, Jishnu, Quy-Toan Do et Berk Ozler, 2005, « A Welfare Analysis of Conditional Cash Transfer Schemes », *World Bank Research Observer*, vol. 20, n° 1, p. 57-80.
- De Janvry, Alain et Elisabeth Sadoulet, 2006, « Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality », *World Bank Economic Review*, vol. 20, n° 1, p. 1-29.
- Fiszbein, Ariel et Norbert Schady, 2009, « Conditional Cash Transfers for Attacking Present and Future Poverty », World Bank Policy Research Report, Banque mondiale, 2009.
- Fujita, Masahisa, Paul Krugman et Anthony Venables, 2001, *The Spatial Economy*, MIT Press, Cambridge, Mass.
- Jacoby, Hanan, 2000, « Access to Markets and the Benefits of Rural Roads », *Economic Journal*, vol. 110, p. 713-737.
- Jalan, Jyotsna et Martin Ravallion, 2002, « Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China », *Journal of Applied Econometrics*, vol. 17, n° 4, p. 329-346.
- King, Elizabeth M. et Jere R. Behrman, 2009, « Timing and Duration of Exposure in Evaluation of Social Programs », *World Bank Research Observer*, vol. 24, n° 1, p. 55-82.
- Klitgaard, Robert, 1988, *Controlling Corruption*, University of California Press, Berkeley CA.
- Knaap, Thijs et Jan Oosterhaven, 2000, « The Welfare Effects of New Infrastructure: An Economic Geography Approach to Evaluating New Dutch Railway Links », miméo, Université Érasme de Rotterdam.
- McCann, Philip, 2005, « Transport Costs and the New Economic Geography », *Journal of Economic Geography*, vol. 5, p. 305-318.
- Mu, Ren et Dominique van de Walle, 2008, « Rural Roads and Local Market Development in Vietnam », Policy Research Working Paper, Banque mondiale.
- Oduor, Jacob, Anne Kamau et Evan Mathenge, 2009, « Evaluating the impact of micro-franchising the distribution of anti-malarial drugs in Kenya on malaria mortality and morbidity », *Journal of Development Effectiveness*, vol. 1, n° 3, septembre 2009, p. 353-377.
- Olken, Benjamin A., 2007, « Monitoring Corruption: Evidence from a Field Experiment in Indonesia », *Journal of Political Economy*, vol. 115, n° 2, p. 200-249.
- Oosterhaven Jan et Thijs Knaap, 2003, Spatial Economic Impacts of Transport Infrastructure Investments, in *Transport projects, programmes, and policies (edited by A. D. Pearman, Peter J. Mackie, John Nellthorp)* Interdisciplinary Centre for Comparative Research in the Social Sciences, Ashgate Publishers.
- Ravallion, Martin, 2008, « Evaluating Anti-Poverty Programs », Paul Schultz et John Strauss (dir. pub.), *Handbook of Development Economics Volume 4*, Amsterdam: North-Holland.

Ravallion, Martin, 2009, « Evaluation in the Practice of Development », *World Bank Research Observer*, vol. 24, n° 1, p. 29-54.

Rogers, Patricia J., 2009, « Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation », *Journal of Development Effectiveness*, vol. 1, n° 3.

Rose-Ackerman, Susan, 1999, *Corruption and Government: Causes, Consequences and Reform*, Cambridge University Press, Cambridge.

Sherlock, Stephen, 2002, « Combating Corruption In Indonesia? The Ombudsman And The Assets Auditing Commission », *Bulletin of Indonesian Economic Studies*, vol. 38, n° 3, p. 367–83.

Skoufias, Emmanuel, 2005, *PROGRESA and Its Impact on the Welfare of Rural Households in Mexico*, Research Report 139, International Food Research Institute, Washington DC.

van de Walle, Dominique, 2009, « Impact Evaluation of Rural Road Projects », *Journal of Development Effectiveness*, vol. 1, n° 1, p. 15-36.

Weiss, Carol, 2001, « Theory-Based Evaluation: Theories of Change for Poverty Reduction Programs », Osvaldo Feinstein et Robert Piccioto (dir. pub.), *Evaluation and Poverty Reduction*, Transaction Publications, New Brunswick, NJ.

White, Howard, 2009, « Theory-based Impact Evaluation: Principles and Practice », *Journal of Development Effectiveness*, vol. 1, n° 3, septembre 2009, p. 271-284.

Pour une adaptation du protocole de l'évaluation d'impact à la nature de l'intervention et à l'objet de l'évaluation

*Patricia Rogers, Professeur d'évaluation du secteur public au CIRCLE (Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation), Royal Melbourne Institute of Technology, Melbourne, Australie.**

Résumé

Un protocole d'évaluation d'impact doit être adapté aux besoins, aux contraintes et aux opportunités du cas considéré, ce que nous appelons l'« adaptation situationnelle » (« *situational responsiveness* »). Le protocole doit correspondre à la nature de l'intervention et aux objectifs de l'évaluation d'impact. L'évaluation doit en particulier aborder les aspects simples, compliqués et complexes de l'intervention. Les aspects simples peuvent être étroitement cadrés et standardisés ; les aspects compliqués agissent dans le cadre d'un ensemble causal ; les aspects complexes sont dynamiques et adaptatifs. Différentes méthodes sont recommandées dans chaque cas – ECR, régression-discontinuité, entretiens non structurés avec la population locale, *Participatory performance story reporting*, ou rapport de performance participatif, et évaluation évolutive.

La démarche d'adaptation situationnelle

Cette session de la conférence visait à démontrer l'application de différentes démarches d'évaluation d'impact. Si les présentations se sont surtout intéressées à la mesure et à l'analyse des causes, il faut garder à l'esprit qu'une évaluation d'impact doit réaliser d'autres tâches (Rogers, 2008a). Un protocole d'évaluation complet indique comment la totalité des tâches en jeu dans une évaluation d'impact seront effectuées – « recensement exhaustif des impacts importants, recueil de données systématique et défendable et analyse des éléments de preuve de ces impacts, inférences raisonnées sur la contribution de l'intervention à la réalisation de ces impacts et gestion efficace de l'évaluation, comprenant une restitution transparente de la méthodologie et, s'il y a lieu, méta-évaluation formelle » (NONIE, Sous-groupe 2, 2008). Diverses options sont possibles pour chacune de ces tâches ; le protocole d'évaluation doit choisir la méthode ou l'association de méthodes la plus appropriée pour chaque tâche.

Certaines approches de l'évaluation d'impact sont centrées sur un protocole ou une méthode particulière de collecte et d'analyse des données ou sur une forme de gouvernance particulière. Mon approche de la conception d'une évaluation d'impact obéit au principe de ce que j'appelle « l'adaptation situationnelle » (Patton, 2008a).

On prend de plus en plus conscience dans le monde du développement comme dans d'autres domaines que les méthodes et techniques appropriées dépendent des situations en jeu. À titre d'exemple, la déclaration du NONIE¹ sur l'évaluation d'impact en 2008 était formulée ainsi :

« Le NONIE défend une approche éclectique et ouverte de la recherche des meilleures méthodes d'évaluation d'impact – qui puise le large éventail de techniques issues de différentes disciplines. » (NONIE, 2008)

Dans son allocution sur l'état des connaissances en matière de mesure de l'efficacité du développement prononcée à la conférence du Groupe indépendant d'évaluation de la

* Contact : Patricia.Rogers@rmit.edu.au 124 Latrobe Street Melbourne VIC 3000 Australie.

Banque mondiale intitulée « *Measuring Development Effectiveness: Progress and Constraints* », Michael Quinn Patton a résumé cette approche conceptuelle comme un recadrage de ce que nous pensons être l'« étalon or » en matière d'évaluation d'impact : « L'étalon or méthodologique n'est pas ici une méthode particulière, mais *l'adéquation* » (Patton, 2008b).

L'adaptation situationnelle consiste à adapter le protocole d'évaluation aux besoins, aux contraintes et aux opportunités de la situation étudiée. Dès lors, les deux questions auxquelles il faut répondre avant d'établir le protocole d'évaluation sont les suivantes : « Quelle est la nature de l'intervention ? » et « Pourquoi réalise-t-on une évaluation d'impact ? »

QUELLE est la nature de l'intervention ?

Le type d'intervention et l'échelle à laquelle elle est réalisée sont des déterminants importants d'un bon protocole d'évaluation. Est-ce un petit projet testé en phase pilote en vue d'une éventuelle reproduction et d'un déploiement à plus grande échelle ? Est-ce un programme en cours qui se poursuivra probablement sous la même forme ? Est-ce une intervention uniforme ou un ensemble d'initiatives disparates ? L'intervention est-elle strictement cadrée et standardisée ou varie-t-elle d'un lieu à l'autre en fonction des conditions, des besoins et des opportunités locaux ? Ces questions ont des implications pour le type d'évaluation d'impact qui sera nécessaire, les ressources dont on pourra probablement disposer pour l'évaluation et les protocoles de recherche possibles.

Quelle est la nature des impacts recherchés ? Sont-ils produits directement par l'intervention ou indirectement ? S'agit-il d'impacts à court terme qui se manifesteront au cours d'un projet et d'une évaluation (tels que les performances scolaires des enfants) ou d'impacts à long terme qui n'apparaîtront que plusieurs années plus tard (tels que l'emploi post-scolaire ou la réhabilitation de sols salinisés) ? S'agit-il d'impacts transformationnels, probablement irréversibles (par exemple apprendre à lire ou à faire du vélo) ou d'impacts fragiles, aisément réversibles (nutrition adaptée ou scolarisation des filles) ? Les impacts peuvent-ils résulter d'une intervention universelle, qui donne des résultats indépendamment du contexte, ou d'une intervention qui ne produit des résultats qu'en présence de circonstances favorables, y compris peut-être d'autres interventions.

Ces caractéristiques permettent d'établir une typologie en trois catégories – simple, compliqué ou complexe (Stacey, 1992 ; Glouberman, 2001 ; Glouberman et Zimmerman, 2002, Kurtz et Snowden, 2003), dont plusieurs travaux ont démontré l'intérêt pour la préparation et l'analyse des évaluations (Guijt, 2008, Patton, 2008a ; Rogers, 2008b). Elle se révèle particulièrement utile pour classer les aspects d'une intervention plutôt qu'une intervention tout entière.

Dans cette typologie, le terme « complexe » revêt une signification particulière et importante, qu'il n'a pas toujours dans l'usage courant. En matière d'évaluation, le terme « complexe » est parfois employé comme synonyme de « compliqué », parfois utilisé pour qualifier toute chose difficile et parfois comme une excuse pour une préparation insuffisante. Ici, « complexe » qualifie des aspects dynamiques et émergents des interventions, qui sont adaptatifs et réactifs aux besoins et opportunités à mesure qu'ils se révèlent. Les aspects simples des interventions peuvent être étroitement prescrits et sont standardisés – c'est le cas par exemple d'un produit, d'une technique ou d'un processus spécifique. Les aspects compliqués d'une intervention comprennent de multiples composantes, s'inscrivent dans une intervention plus vaste comportant plusieurs composantes ou fonctionnent différemment dans le cadre d'un ensemble causal plus vaste, par exemple dans des cadres de mise en œuvre particuliers, pour des types de participants précis ou en association avec une autre intervention.

Ces différents aspects ont d'importantes implications pour le mode de fonctionnement des interventions, la compréhension que nous en avons et l'usage que nous pouvons faire de cette compréhension, comme indiqué au tableau 1.

Tableau 1 Implications des aspects simples, compliqués et complexes des interventions

Aspects	Implications pour :		
	Mise en œuvre de l'intervention	Processus causaux	Restitution et utilisation des constats de l'évaluation d'impact
Simple	Organisation unique	Une seule chaîne causale suffit à produire les impacts	Message unique – ce qui marche
Compliqués (composantes multiples)	Organisations multiples unies par une relation contractuelle avec des fonctions bien définies	De multiples chaînes causales produisent les impacts : Interventions séquentielles multiples ou Interventions simultanées multiples ou Niveaux d'intervention multiples ou Différents mécanismes causaux à l'œuvre dans différents contextes	Message contingent – ce qui marche pour qui, dans quelles situations
Complexes (dynamiques et émergents)	Organisations multiples unies dans une relation de partenariat évolutive	Causalité récursive avec des boucles de rétroaction Résultats émergents – le tout est supérieur à la somme des parties	Message dynamique, émergent – ce qui fonctionne au moment considéré

POURQUOI réalise-t-on une évaluation d'impact ?

L'objectif doit lui aussi être envisagé au moment de la conception du protocole d'évaluation. À quels utilisateurs l'évaluation est-elle destinée ? Que considéreront-ils comme des données crédibles au plan des impacts à considérer, des mesures à effectuer et de l'analyse causale ? Qui doit participer à la sélection des paramètres de l'évaluation ?

Quelles valeurs seront utilisées dans l'évaluation ? Quels impacts, positifs ou négatifs, jugera-t-on significatif ? Quelles distributions des coûts et avantages seront jugées souhaitables ? S'intéressera-t-on à l'effet moyen ou à l'effet sur les plus défavorisés ? Quels usages veut-on faire de l'évaluation ? Est-elle réalisée pour justifier a posteriori les dépenses, auquel cas il suffira d'estimations crédibles des bénéfices nets ? Ou est-elle

réalisée pour guider un éventuel déploiement à grande échelle d'un projet pilote, auquel cas il faudra d'informations de qualité sur ses modes de fonctionnement ? Est-elle destinée à informer un changement incrémentiel ou une restructuration significative d'un programme ou d'une politique ?

Enfin, avant de concevoir une évaluation d'impact, il faut réfléchir aux questions logistiques. À quelle date a-t-on besoin d'un rapport ? De quelles données dispose-t-on déjà sur cette intervention et sur des interventions comparables ? De quelles autres ressources dispose-t-on pour réaliser l'évaluation ?

Dans une démarche d'adaptation situationnelle, ce n'est qu'après avoir réglé toutes ces questions que nous abordons la question « Comment faut-il la réaliser ? ». Bien entendu, en situation réelle, une réponse aurait été apportée à toutes ces questions avant ou pendant l'élaboration du protocole d'évaluation. Ici, nous avons comblé les lacunes des descriptions de cas fin d'envisager des scénarios plus précis.

1. Allocation sociale conditionnelle

Ce cas était décrit comme suit :

« Une allocation sociale conditionnelle dans un pays d'Amérique centrale, versée aux ménages dont les filles d'âge scolaire restent scolarisées et respectent certaines conditions d'assiduité et de résultats. »

Les programmes d'allocations sociales conditionnelles (ASC) présentent d'importants aspects que l'on peut qualifier de simples – c'est-à-dire que ce sont des interventions discrètes, standardisées, qui sont destinées à être déployées à l'identique en différents lieux. Cependant, les évaluations d'ASC qui ont été réalisées ont constaté une hétérogénéité considérable des résultats (par exemple Soares *et al*, 2009). Une ASC peut nécessiter d'autres interventions pour produire les résultats voulus. À titre d'exemple, elle peut permettre de réaliser l'objectif de scolarisation, mais l'amélioration des impacts à plus long terme tels que les acquisitions des élèves, les qualifications et les résultats en termes d'emploi nécessiteront des établissements scolaires efficaces. Il serait donc important de considérer les aspects compliqués de l'intervention et de prévoir de mesurer et d'analyser d'autres composantes (telles que les caractéristiques particulières du cadre de mise en œuvre et des participants) nécessaires à l'obtention des impacts recherchés.

Dans ce cas précis, en supposant que l'objectif de cette évaluation d'impact est de déterminer s'il convient de déployer un programme pilote à grande échelle, un essai contrôlé randomisé (ECR) pourrait être approprié pour l'analyse causale, en particulier s'il était associé à une collecte systématique de données sur d'autres facteurs, tels que la qualité de l'enseignement, qui pourraient être nécessaires pour produire les impacts recherchés, et à une analyse, telle qu'une désagrégation, pour déterminer leur contribution. La qualité de l'analyse serait améliorée en élaborant et en testant une théorie de programme comprenant d'autres facteurs, des effets différentiels et des résultats intermédiaires. Si les ASC permettent effectivement d'améliorer les performances scolaires, mais seulement en association avec des établissements scolaires efficaces, elles n'apporteront pas de solution rapide à elles seules. Puisque la condition d'accès aux ASC est fixée à un certain niveau de revenu, une autre solution pourrait être envisagée, à savoir une méthode de régression-discontinuité, qui pourrait produire des éléments probants de l'attribution causale à moindres frais qu'un ECR, même s'il faudrait aussi la désagréger par environnement de mise en œuvre.

Outre la mesure des impacts recherchés, il est essentiel que l'évaluation d'impact couvre les autres impacts importants, selon la définition donnée par le CAD (Comité d'aide au développement) de l'impact comme les « effets à long terme, positifs et négatifs,

primaires et secondaires, induits par une action de développement, directement ou non, intentionnellement ou non ».

Les impacts non recherchés potentiels peuvent être anticipés et inclus dans la collecte ciblée de données en recourant à une « théorie de programme négative » (Weiss, 1997). Les méthodes participatives accroissent la probabilité d'impacts non anticipés, en particulier d'impacts négatifs (Chambers, 2009). Ensuite, des études de cas, basées sur des entretiens itératifs, des observations et l'étude de documents, pourraient être nécessaires pour faire apparaître les impacts involontaires (positifs ou négatifs), suivis d'enquêtes à grande échelle ou d'examen des statistiques administratives visant à produire des estimations de la fréquence de ces impacts.

2. Programme d'infrastructure de transports

Ce cas était décrit de la manière suivante :

« Évaluation ex-post d'un programme du secteur des transports dans un pays d'Asie du Sud comprenant la réhabilitation d'infrastructures portuaires, la remise en état de routes principales et de nouveaux investissements dans des routes secondaires en zone rurale. »

Nous avons supposé ici que l'objectif de l'évaluation d'impact était d'appréhender l'impact global de l'investissement, principalement en vue d'une restitution aux bailleurs de fonds, mais aussi au public, et qu'aucune évaluation n'avait été prévue au début du programme. Cette carence, alliée à la nature de l'intervention (multi-facettes, diversifiée et affectant une région tout entière), pose des difficultés considérables pour l'analyse causale, car les approches expérimentales et quasi-expérimentales classiques risquent d'être impossibles à déployer efficacement (Ravaillon, 2009).

L'évaluation d'impact doit ici trouver le moyen de recueillir des données sur un grand nombre de composantes diversifiées et de résoudre le problème de l'attribution en l'absence de groupe de comparaison ou de contrôle. Il pourrait donc être opportun ici de recourir à une nouvelle démarche présentée à cette conférence, le *Participatory performance story reporting* (PPSR) ou rapport de performance participatif (Dart, 2009), qui peut exécuter ces fonctions. Puisque pour certaines personnes, le terme « participatif » implique une approche moins rigoureuse du recueil et de l'analyse des données, il est important de comprendre la base empirique très solide qui sous-tend cette approche.

Le PPSR est un développement et une systématisation de la méthode des MLE (*Multiple Lines and Levels of Evidence*), une approche de l'analyse causale conçue pour les situations à forts enjeux où il est impossible de constituer un groupe de contrôle ou un groupe de comparaison, mais où l'on dispose de considérables données de qualité et où une analyse causale systématique est requise.

« Une ligne de données est un type de données, tel qu'un attribut d'un écosystème (par exemple, abondance de poissons, richesse en macro-invertébrés, biomasse des macrophytes), qui est étudié par rapport à un agent stressant ou une intervention ;

Un niveau de données est la valeur de fiabilité à partir de laquelle on infère qu'une activité humaine donnée produit un changement écologique ou géomorphologique. » (Cottingham et al. 2005)

Bien que cette démarche soit absente de la plupart des guides consacrés à l'évaluation d'impact, son utilisation remonte aux stratégies appliquées dans les années 60 dans les analyses de données sur les liens entre le tabagisme et le cancer du poumon. Face à un considérable corpus de données scientifiques, mais en l'absence de données produites par des d'ECR, il était important de pouvoir tirer des conclusions quant aux effets du

tabagisme sur la santé. Plus récemment, la méthode MLLE a été utilisée dans le domaine de l'évaluation des risques humains et écologiques et de la gestion des ressources naturelles (par exemple, Keough *et al*, 2002 ; Boyes, 2006).

La méthode MLLE consiste en une investigation systématique de la force de l'argument causal liant une intervention ou une cause et ses effets, en analysant une association observée en termes de critères de causalité particuliers et en identifiant et en écartant les autres explications possibles. Pour l'association entre le tabagisme et le cancer du poumon, Hill (1965) a considéré sa force, sa cohérence, sa spécificité, sa temporalité, sa cohérence avec d'autres éléments probants acceptés, sa plausibilité et son analogie avec des interventions similaires. Pour investiguer les réponses écologiques à une intervention de gestion des rivières, Norris *et al* (2005) ont considéré la plausibilité biologique, la réaction biologique, la dose et la régularité de l'association. Étant donné le caractère spécialisé et souvent transdisciplinaire des données scientifiques, cette intervention est réalisée par un panel d'experts reconnus d'un ensemble de disciplines, auxquels il est demandé de juger la crédibilité des données et de l'analyse causale.

Le PPSR développe la méthode MLLE pour y ajouter la collecte de données primaires et une réunion des parties prenantes au cours de laquelle les données existantes, les données primaires complémentaires et l'analyse du panel d'experts sont examinées pour produire un rapport d'évaluation d'impact consensuel, qui comprend une trace vérifiable des données probantes (par exemple Clear Horizon *et al*, 2008).

L'application du PPSR au cas relatif à l'infrastructure de transport commencerait par la construction d'une chaîne de résultats, ou théorie du programme, qui identifierait les impacts recherchés et une chaîne de résultats intermédiaires liant le programme à ces impacts. Ensuite, on localiserait les données probantes disponibles sur la mise en œuvre du programme, les résultats intermédiaires obtenus et les impacts finaux, ainsi que la contribution du programme à ces résultats, et on évaluerait leur qualité. Cette évaluation qualitative serait réalisée par rapport au type d'informations et non à une hiérarchie de données basée sur le type de protocole de recherche utilisé. Les entretiens avec des utilisateurs de transports seraient par exemple évalués du point de vue des méthodes d'échantillonnage mises en œuvre, des questions posées, des processus appliqués pour réduire le biais et de l'auditabilité des données. Troisièmement, d'autres données seraient recueillies pour combler des lacunes précises du corpus de données. Par exemple, si les principales données disponibles étaient des statistiques officielles telles que la quantité et le type de marchandises transportées, le nombre de véhicules empruntant les routes et l'activité économique brute, il pourrait être utile de recueillir de brèves descriptions sur l'effet du programme de transport sur les individus, les ménages et les villages. Cette collecte de données complémentaires viserait aussi à recueillir des informations auprès d'un vaste éventail de points de vue sur les effets non recherchés importants, tant positifs que négatifs. Des travaux sur un réseau routier local, par exemple, peuvent faciliter l'accès aux équipements de santé centraux, ce qui améliore le pronostic en cas de complications des accouchements. Des exemples d'impacts résultant d'une petite sélection de ces brèves descriptions pourraient être suivis par un examen des statistiques sanitaires officielles.

Ensuite, un panel d'experts d'un ensemble de disciplines pertinentes serait invité à examiner les données du point de vue de leur aptitude à décrire les impacts et à les attribuer (soit individuellement, soit en association avec d'autres facteurs) au programme. Bien que la méthode PPSR n'établisse pas de lien spécifique entre l'examen des experts et la collecte de données complémentaires, il serait souhaitable que cet examen couvre ces données complémentaires et que le panel recommande de recueillir et d'analyser d'autres données pour combler les lacunes du corpus de données probantes. Enfin, une réunion des représentants des principaux groupes de parties prenantes serait organisée pour examiner le rapport de l'évaluation d'impact. Le rapport final serait publié en différentes versions – un rapport synthétique, reprenant les principaux constats et de brèves descriptions donnant des exemples d'impact, et un rapport détaillé présentant des liens avec l'ensemble des éléments probants sur lesquels

s'appuient les conclusions. Le PPSR permettrait donc probablement de mieux restituer aux diverses parties prenantes, y compris à la communauté, les différents impacts du programme de transport que ne le pourraient d'autres approches.

3. Programme anticorruption

Ce cas était décrit comme suit :

Soutien de donateurs à une commission anticorruption (CAC) dans un pays d'Afrique. Le programme prévoit une aide à l'élaboration de directives, la mise à niveau des infrastructures et des voyages d'étude. Des programmes comparables sont mis en œuvre dans six pays.

Cette intervention n'est pas standardisée et elle ne semble pas non plus faire l'objet de strictes prescriptions préalables. On peut penser en fait que les objectifs du programme, et les moyens de les réaliser, apparaîtront à mesure que le programme sera déployé et qu'on comprendra mieux les priorités et possibilités. C'est pourquoi je proposerais d'utiliser une évaluation « évolutive » (« *developmental evaluation* ») (Patton, 1994, 2008), qui n'est pas destinée à produire un rapport à la fin de la mise en œuvre d'une intervention standardisée et fixe, mais à donner des informations au cours de la mise en œuvre d'une intervention en constante évolution, comportant d'importants aspects complexes.

« L'évaluation évolutive désigne des relations de partenariat à long terme entre des évaluateurs et des acteurs d'initiatives innovantes et du développement. Les processus d'évaluation évolutive supposent de poser des questions à visée évaluative et de recueillir des informations afin de faciliter les décisions évolutives et les corrections au cours du processus. L'évaluateur fait partie d'une équipe dont les membres travaillent ensemble pour conceptualiser, concevoir et tester de nouvelles approches dans un processus constant et durable d'amélioration continue, d'adaptation et de changement intentionnel. La fonction principale de l'évaluateur dans l'équipe est de clarifier des discussions d'équipe par des questions, des données et une logique évaluatives et de faciliter l'évaluation à partir des données relatives à la situation présente, à l'évolution constatée, aux orientations prometteuses, aux pistes à abandonner, aux nouvelles expériences à tenter – autrement dit, des décisions basées sur des données dans le cadre des processus évolutifs de l'innovation. » (Westley *et al*, 2006)

Dans ce cas précis, l'évaluation d'impact passerait par une collaboration avec les personnes chargées de la mise en œuvre du programme et avec les participants afin d'élaborer une théorie du programme correspondant aux problèmes de corruption qu'ils jugent les plus graves et aux actions correctives envisagées. Cette théorie du programme serait construite et revue ultérieurement pour chaque initiative spécifique élaborée dans le cadre du programme, avec des mesures et moyens appropriés pour tester l'attribution causale. Par exemple, si une initiative était ciblée sur une ligne téléphonique permettant au public de signaler anonymement des cas de corruption, l'évaluation d'impact devrait suivre ces cas (en fonction du nombre de signalements et des ressources disponibles, ce suivi pourrait concerner tous les cas ou un échantillon aléatoire stratifié) pour voir ce qui s'est passé du point de vue de l'interlocuteur vers lequel ils ont été orientés, des investigations effectuées, du caractère fondé ou non de la dénonciation et, en cas de dénonciation fondée, des conséquences pour l'auteur des faits reprochés.

En élaborant une théorie du programme pour chaque initiative, il serait possible d'établir différentes théories du changement – par exemple, les interventions étaient-elles destinées à fonctionner en sensibilisant les fonctionnaires aux exigences de probité, en identifiant, sanctionnant et révoquant les fonctionnaires corrompus, en décourageant la corruption en augmentant le risque de détection par des audits formels ou en permettant

au public de signaler les cas de corruption ? Des théories du changement particulières se sont-elles appliquées aux différents types de pratiques de corruption active ou passive dans différents types de programmes ?

Si un objectif prioritaire de l'évaluation était de déterminer les bonnes pratiques et de les reproduire dans d'autres cadres, une approche de « déviance positive » pourrait être efficace. Cette approche suppose d'identifier les sites ou les cas dans lesquels des résultats exceptionnels sont obtenus, d'en vérifier la réalité et d'analyser ce qui produit les résultats. La caractéristique significative de la « déviance positive » est que cette investigation est réalisée par ceux qui cherchent à tirer profit des bonnes pratiques. Elle n'est pas conduite par une équipe d'évaluation, puis diffusée ensuite aux utilisateurs d'informations. La déviance positive a été utilisée dans le domaine de la santé publique, de la nutrition, de l'excision des femmes, de l'éducation et du développement agricole ; un exemple dans le domaine de la corruption a fait l'objet d'un rapport, en l'occurrence des extorsions par des fonctionnaires (Horowitz, 2006).

L'une des difficultés de l'évaluation des programmes anticorruption est l'effet de la corruption elle-même sur la capacité à recueillir des données exactes et à prendre des mesures appropriées à la suite des constats. Dans ce cas, il conviendrait d'être particulièrement attentif à la gestion de l'évaluation afin de garantir son indépendance et la sécurité des investigateurs. On pourrait envisager par exemple d'établir un lien de communication direct avec un niveau de gouvernement digne de confiance afin d'éviter toute oblitération ou altération des constats ou de faire appel à des organisations de défense des citoyens pour recueillir ou investiguer les allégations de corruption.

4. Conclusion

On notera avec intérêt que chacun de ces cas comprendrait une composante d'évaluation basée sur la théorie, même si la théorie du programme serait utilisée de manière très différente. Dans le premier cas, elle sert à déterminer les résultats intermédiaires susceptibles d'indiquer des impacts à plus long terme et à identifier les facteurs contextuels qui doivent être investigués dans l'analyse. Dans le deuxième cas, elle sert de cadre conceptuel pour réunir des éléments probants diversifiés sur un ensemble varié de composantes. Dans le troisième cas, elle sert de cadre conceptuel pour guider un protocole évolutif destiné à recueillir et analyser les données afin d'éclairer le changement en cours.

L'analyse des cas et leurs implications pour l'évaluation d'impact sont résumées au tableau 2 ci-dessous. Bien que ce tableau indique les aspects clés des interventions, il est probable que toutes les interventions comprendraient d'autres aspects à divers degrés (par exemple, tous les cas comporteraient probablement des composantes simples).

Tableau 2 Analyse des caractéristiques et des objectifs des cas étudiés

	Aspects déterminants de l'intervention	Objectifs de l'évaluation
Cas 1 - ASC	Aspects simples – intervention standardisée Aspects compliqués – agit en association avec d'autres facteurs et programmes	Déterminer si ça marche – et dans quels contextes – pour éclairer une politique en cours
Cas 2 – Infrastructure de transport	Complexes – multiples composantes diverses qui doivent fonctionner ensemble efficacement	Évaluer les impacts globaux d'un programme achevé
Cas 3 Anti-corruption	Complexes – intervention adaptative et émergente, adaptée aux besoins, problèmes et opportunités	Comprendre et améliorer un programme évolutif en cours

Les protocoles d'évaluation élaborés pour ces trois cas ont montré que l'adaptation situationnelle suppose de connaître un vaste éventail de méthodes et de techniques. Même si la conduite effective de l'évaluation d'impact devait être sous-traitée à un évaluateur externe possédant une expertise et une expérience appropriées, l'agence délégante aurait besoin d'avoir une connaissance suffisante des méthodes pour pouvoir établir le cahier des charges, sélectionner un consultant approprié, gérer efficacement le contrat, notamment évaluer la qualité du travail. Du point de vue du renforcement des capacités d'évaluation, cela implique que nous avons besoin de profondeur (plus de spécialistes de méthodes particulières) et d'ampleur (comprendre qu'il existe un ensemble de méthodes et quand elles peuvent être les plus appropriées).

Remerciements

Je remercie Michael Patton, Kaye Stevens, Howard White et Bob Williams pour leurs observations judicieuses sur une version antérieure de cet article.

Références

Boyes, B. (2006), « Determining and managing environmental flows for the Shoalhaven River », Report 1 - Environmental Flows Knowledge Review. NSW Department of Natural Resources, mai 2006, téléchargé le 14 mai 2009 à l'adresse : www.dwe.nsw.gov.au/water/pdf/monitor_sholahaven_sh003.pdf

Chambers, R. (2009) « So that the poor count more: using participatory methods for impact evaluation », *Journal of Development Effectiveness*, vol. 1, n° 3, p. 243-246.

Clear Horizon et O'Connor NRM (2008) « Performance Story Report: A study of the Mount Lofty Ranges Southern Emu-Wren and Fleurieu Peninsula Swamps Recovery Program and how it contributed to biodiversity outcomes in the Adelaide and Mount Lofty Natural Resources Management region », Canberra, Australie, téléchargé le 16 mai 2009 à l'adresse : <http://www.nrm.gov.au/publications/books/pubs/psr-mount-lofty.pdf>

Dart, J. (2008) « Report on outcomes and get everyone involved: The Participatory Performance Story Reporting Technique », article présenté lors de la conférence de 2008 de l'Australasian Evaluation Society, Perth, téléchargé le 16 mai 2009 à l'adresse : <http://www.aes.asn.au/conferences/2008/papers/p100.pdf>

Dart, J. (2009) « Participatory Performance Story Reporting Technique », contribution présentée lors de la Conférence de 2009 sur l'évaluation d'impact, le Caire.

Downes, B. J., L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J. Keough, P. S. Lake, B. D. Mapstone et G. P. Quinn (2002), *Monitoring Ecological Impacts: Concepts and Practice in Flowing Waters*, Cambridge University Press, Cambridge.

Glouberman, S. (2001), « Towards a New Perspective on Health Policy », CPRN Study No. H/03, Canadian Policy Research Networks Inc., Ottawa.

Glouberman, S. et B. Zimmerman (2002) « Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? » Commission on the Future of Health Care in Canada, Discussion Paper 8. Téléchargé le 14 mai 2009 sur le site : http://www.hc-sc.gc.ca/english/pdf/romanow/pdfs/8_Glouberman_E.pdf

Guijt, I (2008) « Seeking surprise : rethinking monitoring for collective learning in rural resource management », thèse de doctorat, Wageningen, Pays-Bas.

Hill, A.B. (1965), « The Environment and Disease: Association or Causation », Proceedings of the Royal Society of Medicine, 1965 mai, vol. 58, p. 295-300.

Horowitz, B. (2006), « Bridge Ogres, Little Fishes and Positive Deviants: One-on-one deterrence of Public Functionary Extortion Demands », téléchargé le 14 mai 2009 sur le site : http://www.positivedeviance.org/projects/law/Bridges_final.doc

Kurtz C, D. Snowden (2003), « The new dynamics of strategy: Sense-making in a complex and complicated world », *IBM Systems Journal*, 2003, vol. 42 n° 3, p. 462-483.

Land and Water Australia (), « Improving the Natural Resource Management System for Regions », Canberra, Australian Government Publishing Service, téléchargé le 14 mai 2009 à l'adresse : <http://www.rkrk.net.au/images/3/34/PR061220.pdf>

NONIE Subgroup 2 (Network of Networks on Impact Evaluation) NONIE Impact Evaluation Guidance. Téléchargé le 14 mai 2009 à l'adresse : http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf

NONIE (Network of Networks on Impact Evaluation) Statement on Impact Evaluation. Téléchargé le 14 mai 2009 à l'adresse : http://www.worldbank.org/ieg/nonie/docs/IE_statement.doc

Norris, R., P. Liston, J. Mugodo, S. Nichols (2005), « Multiple Lines and Levels of Evidence for Detecting Ecological Responses to Management Intervention », article présenté à l'American Geophysical Union, Spring Meeting 2005.

Patton, M. Q. (1994), « Developmental Evaluation », *Evaluation Practice*, vol. 15, n° 3, p. 311-319.

Patton, M.Q. (2008a), *Utilization-Focused Evaluation*, 4^e éd., Sage Publications, Thousand Oaks, CA.

Patton, M. Q. (2008b), « State of the Art in Measuring Development Assistance », allocution à la conférence du Groupe d'évaluation indépendante de la Banque mondiale, 10 avril 2008, Washington, DC, téléchargé le 14 mai 2009 à l'adresse : <http://www.worldbank.org/ieg/conference/results/patton.pdf>

Ravaillon, M. (2009), « Evaluating Three Stylized Interventions », *Journal of Development Effectiveness*, vol. 1, n° 3, p. 227-236.

Rogers, P.J. (2008a), « Four key tasks in impact assessment of complex interventions », discours principal, atelier « Rethinking Impact. Understanding the Complexity of Poverty and Change », Consultative Group on International Agricultural Research (CGIAR) Cali-Colombia. Cali, Colombia. Téléchargé le 14 mai 2009 à l'adresse :

<http://www.prgaprogram.org/riw/files/papers/Rogers%20material%20for%20workshop.ppt>

Rogers, P.J. (2008b), « Using programme theory for complicated and complex programmes », *Evaluation: the international journal of theory, research and practice*, vol.14, n° 1, p. 29-48.

Soares, F.V, R. Perez et G.I. Hirata (2009), « Achievements and Shortfalls of Conditional Cash Transfers: Impact Evaluation of Paraguay's Tekoporã Programme », contribution présentée à la Conférence de 2009 sur l'évaluation d'impact, Le Caire.

Stacey, R. (1992), *Managing the Unknowable*, Jossey-Bass, San Francisco.

Weiss, C. (1997), *Evaluation: Methods for Studying Programs and Policies*, (2^e éd.) Prentice Hall, Upper Saddle River, NJ.

Westley, F., B. Zimmerman et M.Q. Patton (2006), *Getting to Maybe: How the World Is Changed*, Random House Canada, extrait téléchargé le 14 mai 2009 à l'adresse : http://innovationlabs.com/r3p_public/rtr3/pre/pre-read/Patton.Developmental%20Evaluation.pdf

NOTE

¹ Le NONIE est un réseau de réseaux pour l'évaluation d'impact, formé du Comité d'aide au développement de l'Organisation de coopération et de développement économiques (CAD-OCDE), d'Evaluation Network, du Groupe des Nations Unies pour l'évaluation (UNEG), de l'Evaluation Cooperation Group (ECG) et de l'Organisation internationale de coopération en évaluation (IOCE), un réseau constitué des associations d'évaluation régionales.