

DIRECTIONS DU DEVELOPPEMENT

Evaluation de l'impact
des projets de Développement
sur la pauvreté

Manuel à l'attention des Praticiens

Judy L. Baker

Banque Mondiale
Washington, D.C.

© 2000 Banque Internationale pour la Reconstruction
et le Développement / BANQUE MONDIALE
1818 H Street, N.W.
Washington, D.C. 20433

Tous droits réservés
Edité aux Etats-Unis d'Amérique
Premier tirage : Mai 2000

Les déclarations, interprétations et conclusions exprimées dans ce document n'engagent que leurs auteurs et ne reflètent aucunement les opinions de la Banque Mondiale ou des institutions qui lui sont affiliées ni du Conseil des Administrateurs ou des pays qu'ils représentent. La Banque Mondiale ne garantit pas l'exactitude des données contenues dans cette publication et n'accepte aucune responsabilité pour toute conséquence découlant de leur usage.

Le contenu de cette publication est protégé par le droit d'auteur. La Banque Mondiale encourage la diffusion de son travail et, dans les conditions normales, accordera rapidement la permission de reproduire des parties de l'ouvrage.

La Banque Mondiale accordera la permission de *photocopier* des articles pour une utilisation interne ou personnelle, pour l'utilisation interne ou personnelle de clients spécifiques ou l'utilisation comme matériel didactique, à condition que les honoraires appropriés soient payés directement à Copyright Clearance Center, Inc. (Bureau du Droit d'Auteur), 222 Rosewood Drive, Danvers, MA 01923, USA ; téléphone 978-750-8400, fax 978-750-4470. Prière de contacter le Bureau du Droit d'Auteur avant la photocopie des articles.

Pour obtenir la permission de réimprimer des articles ou des chapitres, prière d'envoyer un fax de votre demande avec information complète au Département de la Réédition, Copyright Clearance Center, Inc., Fax 978-750-4470.

Pour plus de renseignements sur les droits et licences, prière de vous adresser au Bureau de l'Editeur, Banque mondiale, à l'adresse ci-dessus ou envoyer un fax au 202-522-2422.

ISBN

0-8213-4697-0

Bibliothèque du Congrès. Fichier des données de la publication

Baker, Judy L., 1960-

Evaluation de l'impact des projets de développement sur la pauvreté: Manuel à l'Attention des Praticiens / Baker, Judy L.,
p. cm. – (Directions du développement)
Références bibliographiques incluses.
ISBN 0-8213-4697-0

1. Projets de développement économique –Evaluation – Guide, manuels etc.
2. Pays pauvres en développement I. Titre, II. Direction du développement (Washington, D.C.)

HD75.9 .B35 2000

338.9'0068'4 – dc21

00-028325

Table des matières

Avant-propos	vi
Remerciements	viii
1 Définition des concepts et techniques d'évaluation d'impact	1
2 Démarches clés dans la Conception et la mise en œuvre des évaluations de l'impact	12
3 Application des méthodes analytiques d'évaluation de l'impact : Une étude de cas	30
4 Exploiter les évaluations de l'impact de la « Meilleure Pratique »	48
5 Bibliographie	62
6 Annexes	73
Annexes 1 : Etudes de cas	73
1.1 Evaluation des acquis des pauvres du programme protection-travail (programme associant aide sociale et aide à l'emploi) Le programme TRABAJAR de l'Argentine	73
1.2 La micro-finance aide-t-elle vraiment les pauvres ? Preuves nouvelles à partir des programmes les plus importants au Bangladesh	79

1.3 Education contre vivres au Bangladesh : Evaluation d'un programme social ciblé lorsque le placement est décentralisé	105
1.4 Evaluation du fonds d'investissement social de la Bolivie	109
1.5 Impact des programmes de travail actifs : République Tchèque	114
1.6 Impact du crédit et éducation sur la nutrition des mères et de leurs jeunes enfants : Programme de la Banque rurale du « Lower Pra » au Ghana	119
1.7 Manuels scolaires et résultats des tests : Cas concret d'une éventuelle évaluation au Kenya	123
1.8 Evaluation du projet de vulgarisation agricole au Kenya	128
1.9 Impact du Programme de Recyclage du Mexique sur l'emploi et les salaires (PROBECAT)	134
1.10 Programme national pour l'éducation, la santé, et la nutrition (PROGRESA) au Mexique – (Proposition d'évaluation)	140
1.11 Evaluation de la Réforme scolaire au Nicaragua : une approche quantitative et qualitative combinée	145
1.12 Améliorer l'enseignement des mathématiques élémentaires au Nicaragua : Une étude expérimentale de l'impact des manuels et de la radio sur la performance	151
1.13 Impact des programmes alternatifs de recouvrement des coûts sur l'accès aux soins de santé et l'équité au Niger	156
1.14 Résultats scolaires dans les établissements primaires aux Philippines : Evaluation de l'impact à partir de quatre expériences	160
1.15 Evaluation de l'impact des projets de routes rurales sur la pauvreté au Vietnam	165
Annexe 2 : Modèle de termes de référence	169
2.1 Projet de nutrition et de développement de la petite enfance de l'Ouganda	169
2.2 Evaluation de l'impact des routes rurales : Niveau de référence de 1997 au Vietnam	188
Annexe 3 : Budget type d'une évaluation de l'impact d'un Programme alimentaire au niveau scolaire	195
Annexe 4 : Indicateurs d'impact – Evaluation du fond social d'investissement en Bolivie	198
Annexe 5 : Modèle de registre des récapitulatifs de conception de projet pour le document d'achèvement du projet ou document d'évaluation du projet	204
Annexe 6 : Matrice d'analyse – Evaluation de l'impact du fond social d'investissement d'urgence au Nicaragua	208

TABLES DES MATIERES V

Encadrés

1.1 Le problème de la distorsion de sélection	5
1.2 Résumé des méthodes quantitatives pour évaluer l'impact de programme	6
1.3 Résumé des méthodes employées pour évaluer des politiques d'ajustement structurel	11
2.1 Principales démarches dans la conception et l'exécution d'évaluations d'impact	17
2.2 Points clés pour identifier des ressources de données pour l'évaluation d'impact	21
3.1 Etapes de l'harmonisation du score de propension	50
3.2 Sources de distorsion dans les estimations primaires de l'impact du PROSCOL	53
3.3 Faire une différence double	56
3.4 Mesures de la pauvreté	59
3.5 Comparaison de la pauvreté avec ou sans le Programme	60

Tableaux

2.1 Contrôle des méthodologies d'évaluation et concordance des données	28
2.2 Principaux instruments de collecte de données pour l'évaluation d'impact	32
4.1 Résumé des évaluations d'impact de la « Meilleure pratique »	67
4.2 Résumé des coûts estimés à partir de plusieurs évaluations d'impact de la Banque Mondiale	79

Avant-Propos

Malgré les milliards de dollars injectés dans l'aide au développement chaque année, nous savons encore peu de chose sur l'impact réel des projets sur les pauvres. Les bénéfices de la croissance économique, les investissements en capital humain et la mise à disposition de filets de sécurité pour les pauvres sont largement documentés. Cependant, pour un programme ou projet spécifique d'un pays donné, l'intervention produit-elle les bénéfices escomptés et quel a été l'impact global sur la population ? Le programme ou projet pourrait-il être mieux conçu pour atteindre les résultats escomptés ? Les ressources sont-elles dépensées de façon efficiente ? Tels sont les types de questions auxquelles on ne peut répondre que par une évaluation d'impact, une approche qui mesure les résultats d'une intervention de programme en éliminant les autres facteurs possibles.

Beaucoup de gouvernements, d'institutions et de directeurs de projet sont réticents à effectuer des évaluations d'impact parce qu'on les considère trop coûteuses, prenant du temps et techniquement complexes et parce que les résultats peuvent être politiquement sensibles, surtout s'ils sont négatifs. Beaucoup d'évaluations ont aussi été critiquées parce que les résultats arrivent trop tard, ne répondent pas aux bonnes questions ou n'ont pas été effectués avec la rigueur analytique suffisante. Une contrainte supplémentaire est souvent la disponibilité limitée et la qualité des données.

Pourtant avec une planification appropriée et précoce, l'appui des décideurs et un investissement relativement minime comparé au coût global du projet, une évaluation rigoureuse peut avoir une très forte influence sur l'estimation du caractère approprié et efficient des programmes.

L'évaluation de l'impact est particulièrement critique dans des pays en développement où les ressources sont rares et chaque dollar dépensé doit aspirer à maximiser son impact sur la réduction de la pauvreté. Si les programmes sont mal conçus, n'atteignent pas les populations cibles ou sont peu rentables, avec l'information juste ils peuvent être reformulés, améliorés, ou le cas échéant abandonnés. La connaissance acquise à partir des études d'évaluation d'impact fournira aussi la ressource décisive pour la conception appropriée de programmes et de projets futurs.

Ce manuel cherche à fournir aux directeurs de projet et aux analystes de politique les outils nécessaires pour évaluer l'impact d'un projet. Il est destiné aux

Avant-Propos

lecteurs qui ont une connaissance générale des statistiques. Pour certaines des méthodes statistiques traitées plus en profondeur le lecteur est prié de se reporter à la littérature technique sur le sujet. Le chapitre 1 présente une vue d'ensemble des concepts et des méthodologies, le chapitre 2 traite des démarches clés et des questions connexes à considérer dans la mise en œuvre, le chapitre 3 illustre les diverses techniques analytiques d'évaluation par une étude de cas et le chapitre 4 contient une discussion de leçons tirées d'un riche jeu d'évaluations de " Meilleure Pratique" de projets de pauvreté qui ont été revues pour ce manuel. Les études de cas contenues dans l'annexe 1 ont été sélectionnées à partir d'une gamme d'évaluations effectuées par la Banque Mondiale, d'autres agences de donateurs, des institutions de recherche et des cabinets de conseil privés. Ils ont été choisis en fonction de leur rigueur méthodologique, dans une tentative de couvrir un éventail varié de situations de pays, de types de projets et de méthodologies d'évaluation.

Les annexes comportent également des échantillons de composantes essentielles qui seraient nécessaires pour planifier n'importe quelle évaluation d'impact – des modèles de termes de référence, budgets, indicateurs d'impact, cadre de conception et matrice d'analyse.

Bien que les techniques employées dans l'évaluation d'impact soient semblables à travers des secteurs et des sous-groupes de population, les illustrations de méthodologies et les exemples de cas dans ce manuel se concentrent sur l'évaluation de l'impact sur des projets axés sur les pauvres. L'impact sur la pauvreté peut englober une vaste gamme de projets et de questions d'évaluation, comme la mesure de l'impact de programmes de micro-finance sur le revenu du ménage, l'impact d'un programme de formation sur l'emploi, l'impact d'un programme de cantine scolaire, sur l'assiduité de l'étudiant ou l'impact de la reconstruction de routes rurales sur le bien-être du ménage.

Indépendamment du type de projet ou des questions traitées, la conception de chaque évaluation d'impact sera unique, selon des facteurs comme le type de données disponibles, les capacités locales et les préoccupations de programmation du temps et de budget y afférents. Finalement, les évaluations généreront des résultats de haute qualité, crédibles et généralisables pour des décideurs exigeront, un appui financier et politique fort ; une planification précoce et prudente, une participation des partenaires dans la conception des objectifs et l'approche de l'étude, des données fiables ; une combinaison appropriée de méthodologies, y compris des techniques tant quantitatives que qualitatives ; l'application rigoureuse de ces techniques ; et la communication entre les membres d'équipe tout au long du processus.

Remerciements

La préparation de ce livre a bénéficié des contributions inestimables d'une équipe clé. Je voudrais remercier, tant pour leurs contributions écrites que pour leurs commentaires utiles au cours du travail, les membres d'équipe suivants : Gillette Hall (études de cas, leçons tirées), Julia Lane (études de cas, leçons tirées), Martin Ravallion (étude de cas de méthode analytique), et Laura Rawlings (questions de la mise en œuvre, leçons tirées), et le travail sur l'évaluation d'impact effectué par Kene Ezemenari, Gloria Rubio, Anders Rudqvist, et K. Subbarao. La recherche de fond a été effectuée par Matthew Fleming et Samir Stewart. Le livre a bénéficié du soutien conjoint de la Région Amérique Latine et Caraïbes et du Réseau pour la Réduction de la pauvreté et de la Gestion Economique de la Banque mondiale sous la direction de Norman Hicks, Guillermo Perry et Michael Walton. Le travail a aussi bénéficié énormément des commentaires reçus de Omar Arias, Sabina Alkire, Michael Bamberger, Soniya Carvalho, Wendy Cunningham, Norman Hicks, Shahidur Khandker, Norbert Schady, et Quentin Wodon.

Chapitre 1

Définition des Concepts et Techniques pour l'Évaluation d'Impact

Une évaluation globale se définit dans la littérature comme une évaluation qui intègre le contrôle, l'évaluation du processus, l'évaluation du coût-bénéfice et l'évaluation d'impact. Cependant chacune de ces composantes est nettement différente. Le contrôle aidera à évaluer si un programme est mis en œuvre comme il a été planifié. Un système de contrôle de programme permet la rétroaction continue sur l'état de la mise en œuvre du programme, l'identification des problèmes spécifiques comme ils apparaissent. . L'évaluation du processus traite du fonctionnement du programme et se concentre sur des problèmes de prestation de service. Les évaluations coût-bénéfice ou de rentabilité évaluent des coûts de programme (monétaire ou non monétaire), en particulier leur relation avec des utilisations alternatives des mêmes ressources et des bénéfiques produits par le programme. Et finalement, l'évaluation d'impact est destinée à déterminer de façon plus large si le programme a eu l'impact désiré sur des individus, des ménages et des institutions et si ces effets sont attribuables à l'intervention du programme. Les évaluations d'impact peuvent aussi explorer des conséquences imprévues, soit positives soit négatives sur les bénéficiaires.

L'intérêt particulier de ce manuel est la mesure dans laquelle les bénéfiques du projet atteignent les pauvres et l'impact que ces bénéfiques ont sur leur bien-être. Certaines des questions traitées dans l'évaluation d'impact sont entre autres : quel a été l'impact du projet sur les bénéficiaires ? Certaines améliorations ont –elles résulté directement du projet ou bien auraient-elles de toute façon été obtenues ? La conception du programme pouvait-elle être modifiée pour améliorer l'impact ? Les dépenses étaient-elles justifiées ?

Ces questions ne peuvent pas, cependant, être simplement évaluées par le résultat d'un projet. D'autres facteurs ou événements peuvent être en corrélation avec les résultats, sans être engendrés par le projet. Pour assurer la rigueur méthodologique une évaluation d'impact doit estimer les effets contre factuels c'est-à-dire ce qui serait arrivé si le projet n'avait jamais vu le jour ou ce qui aurait pu arriver. Par exemple, si un diplômé récent d'un programme de formation de travail trouve un emploi, est-ce que c'est un résultat direct du programme ou est-ce que l'individu de toute façon aurait trouvé du travail ? Pour déterminer la comparaison contre factuelle il est nécessaire de distinguer l'effet des interventions des autres facteurs – une tâche quelque peu complexe. Cela se fait à travers la comparaison de groupes témoins (ceux qui ne participent pas à un programme ou n'en reçoivent pas de bénéfiques) avec les groupes cibles (les individus qui reçoivent l'intervention). Les groupes

témoins sont choisis aléatoirement de la même population que les participants au programme tandis que le groupe de comparaison est tout simplement le groupe qui n'est pas pris en compte par l'étude du Programme. Les groupes de comparaison aussi bien que les groupes de contrôle devraient ressembler au groupe cible en tous points, la seule différence entre les groupes étant la participation au programme.

La détermination de la comparaison contre factuelle est au cœur de la conception d'évaluation. Elle peut être réalisée selon plusieurs méthodologies regroupées en deux grandes catégories, des conceptions expérimentales (aléatoires) et des conceptions quasi expérimentales (non aléatoires). Il est cependant difficile de distinguer l'impact du programme des conditions contre factuelles qui peuvent être affectées par l'histoire, la distorsion de la sélection et la contamination. Des méthodes qualitatives et participatives peuvent aussi être utilisées pour évaluer l'impact. Ces techniques présentent souvent des vues critiques sur les perspectives des bénéficiaires, la valeur des programmes pour les bénéficiaires, les processus qui auraient pu affecter les résultats et une interprétation plus approfondie des résultats observés à l'analyse quantitative. Les forces et les faiblesses de chacune de ces méthodes sont traitées plus en détail ci-dessous. Comme le lecteur le constatera, aucune technique n'est parfaite et par conséquent l'évaluateur doit décider quant aux options pour chaque méthode choisie. La planification précoce et soignée présentera cependant beaucoup plus d'options méthodologiques dans la conception de l'évaluation.

Conceptions Expérimentales

Les conceptions expérimentales également connues comme aléatoires sont généralement considérées comme la plus solide des méthodes d'évaluation. En affectant de manière aléatoire l'intervention parmi les bénéficiaires éligibles, le processus de nomination lui-même crée les groupes cibles et les groupes témoins comparables qui sont statistiquement équivalents les uns des autres étant donné les tailles d'échantillon appropriées. C'est un résultat très positif parce que, dans la théorie les groupes témoins produits par la nomination aléatoire servent comme une parfaite comparaison contre factuelle libre des questions de distorsion de sélections incommodes qui existent dans toutes les évaluations. L'avantage principal de cette technique est la simplicité dans l'interprétation des résultats – l'impact du programme sur le résultat que l'on évalue par la différence entre les moyennes des échantillons du groupe cible et le groupe témoin. Un exemple est l'évaluation de manuels scolaires au Kenya dans laquelle les évaluateurs ont sélectionné une répartition aléatoire de sites de programme, ont mené une enquête de base, créé des groupes de contrôle et ont ensuite administré le traitement qui dans ce cas était la fourniture de manuels scolaires. Le fait d'avoir des groupes de contrôle et des groupes cibles a alors permis aux évaluateurs de déterminer clairement l'impact des manuels sur l'éducation de l'étudiant.

Tandis que l'on considère les conceptions expérimentales comme l'approche optimale à l'évaluation de l'impact d'un projet, en pratique plusieurs problèmes subsistent. D'abord la simulation de distribution aléatoire peut être contraire à la morale par suite du refus des bénéficiaires ou services

aux membres autrement éligibles de la population pour les objectifs de l'étude. Un exemple extrême serait le refus de traitement médical qui peut s'avérer être d'importance vitale pour certaines tranches d'une population. Deuxièmement, il peut être politiquement difficile de fournir une intervention à un groupe et pas à un autre. Troisièmement, la portée du programme peut signifier qu'il n'y a pas de groupes de non-cible en tant que tels, avec un projet ou un changement de politique qui a une large portée – par exemples un prêt d'ajustement ou des programmes administrés à un niveau national. Quatrièmement, des individus au sein des groupes témoins peuvent changer certaines caractéristiques d'identification pendant l'expérience, ce qui pourrait infirmer ou affecter les résultats. Si, par exemple, les gens se déplacent à l'intérieur et à l'extérieur d'un secteur de projet, ils peuvent se déplacer à l'intérieur et à l'extérieur du groupe cible ou du groupe témoin. Alternativement les gens à qui on a refusé un avantage de programme peuvent le chercher par des sources alternatives ou ceux à qui un programme est destiné peuvent ne pas jouir de l'intervention. Cinquièmement, il peut être difficile d'assurer que la sélection est vraiment aléatoire. L'exemple pourrait être les administrateurs qui excluent des candidats à haut risque pour réaliser de meilleurs résultats. Et finalement des conceptions expérimentales peuvent être coûteuses et prendre beaucoup de temps dans certaines situations, particulièrement dans la collecte de nouvelles données.

La planification rigoureuse peut permettre de surmonter certains de ces problèmes dans la mise en œuvre de conceptions expérimentales. La sélection aléatoire des bénéficiaires constitue une voie. Cela peut être employé pour fournir à la fois un mécanisme d'attribution politiquement transparent et la base d'une conception d'évaluation saine comme le budget où des contraintes de l'information rendent souvent impossible d'identifier exactement et d'atteindre les bénéficiaires les plus éligibles. Une deuxième voie se propose d'inclure des groupes témoins dans le programme à une étape postérieure une fois que l'évaluation a été conçue et amorcée. Dans cette technique, le choix aléatoire détermine plutôt quand le bénéficiaire éligible reçoit le programme ou s'il ne le reçoit pas. Cette technique a été appliquée dans l'évaluation d'un programme de nutrition en Colombie qui a présenté l'avantage complémentaire de traiter des questions relatives au temps nécessaire imparti pour que le programme devienne efficace dans la réduction de la sous alimentation (McKay 1978). Finalement la simulation de distribution aléatoire peut être appliquée dans un sous ensemble de bénéficiaires également éligibles, en atteignant tous les candidats les plus éligibles et en refusant les bénéficiaires au moins éligibles comme cela a été fait pour l'évaluation du fonds social des projets d'éducation dans la région d'El Chaco en Bolivie (Pradhan, Rawlings et Ridder 1998). Cependant, si la dernière suggestion est mise en œuvre, il faut garder à l'esprit que les résultats de l'évaluation seront applicables au groupe dont l'échantillon aléatoirement constitué a été choisi.

Conceptions Quasi Expérimentales

Des méthodes quasi expérimentales (non aléatoires) peuvent être employées pour effectuer une évaluation quand il n'est pas possible de constituer les groupes cible et de comparaison

par une conception expérimentale. Ces techniques produisent les groupes de comparaison qui ressemblent au groupe cible, au moins dans des caractéristiques observées, par des méthodologies d'économétrie, qui comprennent des méthodes correspondantes, des méthodes de différence doubles, des méthodes de variables instrumentales et des comparaisons réflexives (voir l'Encadré 1.) Quand ces techniques sont employées, les groupes de traitement et les groupes de comparaison sont d'habitude choisis après l'intervention en employant des méthodes non aléatoires. Donc, des contrôles statistiques doivent être appliqués pour traiter des différences entre les groupes cibles et les groupes de comparaison et des techniques de correspondance sophistiquées doivent être utilisés pour constituer un groupe de comparaison qui est aussi semblable que possible au groupe cible. Dans certains cas un groupe de comparaison est aussi sélectionné avant le traitement, bien que la sélection ne soit pas aléatoire.

L'avantage principal des conceptions quasi-expérimentales est qu'elles peuvent partir de sources de données existantes et sont ainsi souvent plus rapides et peu coûteuses à mettre en œuvre et elles peuvent être exécutées après qu'un programme ait été mis en œuvre, compte tenu des données existantes suffisantes. Les principaux inconvénients des techniques quasi-expérimentales sont : (a) la fiabilité des résultats est souvent réduite puisque la méthodologie est statistiquement moins rigoureuse ; (b) les méthodes peuvent être statistiquement complexes ; et (c) il y a un problème de sélection biaisée. En créant un groupe de comparaison plutôt que d'en choisir un aléatoirement, beaucoup de facteurs peuvent affecter la fiabilité des résultats. La complexité statistique exige l'expertise considérable dans la conception de l'évaluation et dans l'analyse et l'interprétation des résultats. Cela ne peut pas toujours être possible, particulièrement dans certaines circonstances des pays en voie de développement.

Le troisième problème de distorsion de choix tient au fait que les sous-groupes d'une population cible participe différemment à un programme, affectant ainsi l'échantillon et en fin de compte les résultats. Il y a deux types de sélection biaisée : ceux en raison des différences observables ou autre dans les données et ceux en raison des différences non observables (pas dans les données), souvent appelées sélection biaisée (Encadré 1.1) Une sélection biaisée observable pourrait comprendre les critères de choix par lesquels un individu est identifié, comme l'emplacement géographique, la scolarisation ou la participation au marché du travail. Les éléments non observables qui peuvent biaiser les résultats de programme pourraient inclure la capacité individuelle, l'empressement à travailler, des rapports de famille et un processus subjectif de choix (souvent politiquement orienté) des individus pour un programme. Les deux types de distorsion de sélection peuvent engendrer des résultats imprécis, comprenant une sous estimation et surestimation d'impacts réels de programme d'impacts négatifs, quand les impacts réels de programme sont positifs (et vice versa) et des impacts statistiquement insignifiants quand des impacts réels de programme sont significatifs et vice versa. (Voir, par exemple, LaLonde 1986, Fraker et Maynard 1987, Lalonde et Maynard 1987 et Friedlander et 1995 Robins.) Il est possible de maîtriser le biais par des techniques statistiques comme la combinaison, et les variables instrumentales, mais il est très difficile de les corriger entièrement

ce qui reste un défi principal pour des chercheurs dans le domaine d'analyse d'impact.

Parmi les techniques de conception quasi-expérimentale, les techniques de combinaisons mixtes sont généralement considérées comme une seconde meilleure alternative de la conception expérimentale. La majorité de la littérature sur la méthodologie d'évaluation est centrée sur l'utilisation de ce type d'évaluation, reflétant tant la fréquence d'utilisation des groupes de comparaison que les nombreux défis posés par le fait d'avoir des groupes de comparaison moins meilleurs. Ces dernières années il y a eu des avancées substantielles dans la propension des techniques de combinaisons de score. (Rosenbaum et Rubin 1985 ; Jalan et Ravallion 1998). Cette méthode est

Encadré 1.1 Le Problème de la Distorsion de Choix

La sélection biaisée est relative aux éléments inobservables qui peuvent déformer les résultats (par exemple la capacité individuelle, les conditions préexistantes). Des expériences aléatoires résolvent le problème de la distorsion de choix en créant un groupe témoin expérimental de gens qui auraient participé à un programme mais à qui on avait délibérément refusé l'accès au programme ou le traitement. La nomination aléatoire n'enlève pas la sélection biaisée mais équilibre par contre la distorsion entre les échantillons participants et non participants. Dans les conceptions quasi-expérimentales, les modèles statistiques (par exemple, la concordance, les différences doubles des variables instrumentales) abordent ceci en modélant les processus de sélection pour arriver à une estimation non distorsion utilisant des données non expérimentales. L'idée générale consiste à comparer des participants du programme et les non participants en maintenant des processus de sélection constants. La validité de ce modèle est tributaire de la spécification du modèle.

Les salaires des femmes constituent un bon exemple. Les données représentent les femmes qui veulent travailler. Si cette décision était prise, nous pourrions ignorer le fait que ce ne sont pas tous les salaires qui sont évalués et utiliser la régression ordinaire pour évaluer un modèle de salaire. Certes, la décision des femmes de travailler n'est pas prise au hasard – des femmes qui auraient de bas salaires ne choisiraient vraisemblablement pas de travailler parce que leur salaire minimum accepté est plus élevé que le salaire offert par les employeurs. Ainsi l'échantillon de salaires observés pour des femmes serait tiré vers le haut.

Cela peut être corrigé s'il y a des variables qui affectent fortement les chances pour l'observation (le salaire minimum accepté) ; mais pas le résultat de l'étude (le salaire offert). Une telle variable pourrait être le nombre d'enfants à charge à la maison.

Source : Greene (1997).

très attirante pour les évaluateurs qui ont des contraintes de temps et qui travaillent sans le bénéfice de données de base étant donné qu'elles peuvent être utilisées avec un échantillon de données. Cette technique est cependant tributaire de l'acquisition de données fiables parce qu'elle repose sur des bénéficiaires d'un programme sur-échantillonné pendant la suspension d'une plus large enquête en les associant ensuite à un groupe de comparaison choisi de plus grand échantillon de l'effort d'ensemble, souvent une enquête nationale sur les ménages. Compte tenu de la croissance des demandes de grandes enquêtes dans les pays en voie de développement, tel que les Etudes de Mesure de niveaux de vie à objectifs multiples, cette méthode d'évaluation semble être particulièrement prometteuse. Un bon exemple est l'évaluation d'un programme de travaux publics -TRABAJAR- en Argentine (Jalan et 1998 Ravallion, l'Annexe 1.1 et le chapitre 4).

Encadré 1.2 Résumé des Méthodes Quantitatives pour Evaluer l'Impact de Programme

Les principales méthodes pour l'évaluation d'impact sont traitées ci-dessous. Etant donné, qu'aucune méthode n'est parfaite, il est toujours souhaitable de procéder par triangulation

Conception de Vérification Expérimentale ou Aléatoire.

- La simulation aléatoire dans laquelle le choix du traitement et des groupes témoins est aléatoire au sein de quelques groupes de gens bien définis. Dans ce cas il ne doit y avoir aucune différence (dans l'attente) entre les deux groupes en plus du fait que le groupe cible avait accès au programme. (Il peut toujours y avoir des différences en raison d'erreur d'échantillonnage ; plus la taille des échantillons de traitement et de vérification est grande, moins il y a d'erreur).

Des conceptions non expérimentales ou quasi-expérimentales

- Harmonisation des méthodes ou vérifications construites dans lesquelles on essaye de choisir des comparaisons idéales qui correspondent au groupe de traitement d'une enquête généralisée. Le type d'harmonisation le plus employé est l'harmonisation du score de propension, dans laquelle le groupe de référence est mis en relation avec le groupe de traitement sur la base d'un jeu de caractéristiques observées ou en employant "le score de propension" (la probabilité de la participation prévue compte tenu des caractéristiques observées); plus le score de propension se rapproche, et meilleure est l'harmonisation. Un bon groupe de comparaison

vient du même environnement économique et était soumis au même questionnaire par des enquêteurs formés de la même manière que le groupe de traitement.

- *Méthodes de la double différence ou de la divergence dans la différence*, dans lesquelles on *compare* un groupe de traitement et un groupe de comparaison (la première différence) avant et après un programme (la deuxième différence). Ceux qui opèrent la comparaison devraient être arrêtés quand ils utilisent les scores de propension et qu'ils ont des scores qui se situent en dehors de la gamme observée pour le groupe de traitement.
- *Des variables instrumentales ou des méthodes de vérification statistiques*, dans lesquelles on emploie une ou plusieurs variables qui concernent la participation, et non les résultats de la participation donnée. Cela identifie la variation exogène dans les résultats attribuables au programme, reconnaissant que son placement n'est pas aléatoire, mais intentionnel. «Les variables instrumentales» sont d'abord utilisées pour prévoir la participation au programme ; alors on voit comment l'indicateur de résultat varie avec les valeurs prévues.
- *Des comparaisons réflexives*, dans lesquelles une enquête initiale est faite sur des participants avant l'intervention et une enquête complémentaire est faite après. L'enquête de base fournit le groupe de comparaison et l'impact est mesuré par le changement d'indicateur de résultat avant et après l'intervention.

Méthodes Qualitatives

Des techniques qualitatives sont aussi utilisées pour faire l'évaluation d'impact dans l'intention de déterminer l'impact par la confiance à quelque chose d'autre que la comparaison avec un groupe témoin pour faire une interférence causale (Mohr 1995). Le centre d'intérêt se situe en revanche sur la compréhension du processus, des comportements et conditions tels qu'ils sont perçus par les individus ou les groupes étudiés (Valadez et Bamberger 1994). Par exemple, l'observation des méthodes qualitatives et particulièrement l'observation des participants peut permettre de comprendre les façons dont les ménages et les communautés locales perçoivent un projet et comment celui-ci les affecte. Parce que la mesure de la comparaison avec un groupe témoin se situe au cœur des techniques d'analyse d'impact, des conceptions qualitatives ont été généralement employées conjointement avec d'autres techniques d'évaluation. L'approche qualitative emploie des méthodes relativement ouvertes pendant la conception, la collecte des données et l'analyse. Des données qualitatives peuvent aussi être quantifiées. Parmi les méthodologies employées dans des évaluations d'impact qualitatives, il y a les techniques développées pour l'évaluation rurale rapide, qui repose sur la connaissance par les participants des conditions environnantes du projet ou du programme dans toutes les étapes de l'évaluation ou des évaluations

participatives dans lesquelles les partenaires sont impliqués dans toutes les étapes de l'évaluation - détermination de l'évaluation des objectifs de l'étude, identification et sélection des indicateurs à utiliser et participation à la collecte de données et à l'analyse. Pour une discussion détaillée des méthodes participatives voir la Banque Mondiale (1996), *Le "Livre de Base sur l'évaluation participative de la Banque Mondiale*.

Les avantages des évaluations qualitatives consistent en ce qu'elles sont flexibles, peuvent être spécifiquement construites aux besoins de l'évaluation utilisant des approches ouvertes, peuvent être effectuées rapidement en utilisant des techniques rapides et peuvent énormément accroître les résultats d'une évaluation d'impact en fournissant une meilleure compréhension des perceptions et priorités des partenaires et des conditions et processus qui auraient pu affecter l'impact du programme.

Parmi les inconvénients majeurs, il y a la subjectivité en jeu dans la collecte des données, le manque d'un groupe de comparaison et le manque de rigueur statistique étant donné les tailles principalement petites des échantillons, rendant difficile la généralisation à une population plus large et représentative. La validité et la fiabilité des données qualitatives dépendent fortement de la compétence méthodologique, la sensibilité et la formation de l'évaluateur. Si le personnel de terrain n'est pas sensible aux normes et pratiques sociales et culturelles spécifiques et aux messages non verbaux, les données collectées peuvent être mal interprétées. Et finalement sans un groupe de comparaison, il est impossible de déterminer le groupe témoin et ainsi la conséquence d'impact du projet.

Combinaison de Méthodes Quantitatives et Qualitatives

Bien qu'il existe une large littérature sur les méthodes qualitatives et quantitatives d'impact, il y a aussi une acceptation croissante du besoin de combiner les deux approches. Les évaluations d'impact utilisant des données quantitatives d'échantillons statistiquement représentatives sont meilleures et conviennent à l'évaluation de la causalité utilisant des méthodes économétriques ou à l'obtention d'une conclusion généralisable. Cependant, des méthodes qualitatives permettent l'étude approfondie des questions sélectionnées, des cas ou des événements et peuvent présenter des aperçus critiques dans les perceptions des bénéficiaires la dynamique d'une réforme particulière ou les causes cachées de certains résultats observés dans une analyse quantitative. Il y a des différences significatives dans le choix d'une technique au lieu d'une autre.

La combinaison des évaluations quantitatives et qualitatives peut souvent s'avérer la meilleure voie pour répondre aux besoins d'information du projet. Dans la fusion des deux approches, des méthodes qualitatives peuvent être employées pour informer sur les questions clés de l'évaluation d'impact, examiner le questionnaire ou la stratification de l'échantillon qualitatif et analyser le contexte social, économique et politique dans lequel un projet se situe tandis que des méthodes quantitatives peuvent être employées pour informer des stratégies de collecte de données qualitatives concevoir l'échantillon, informer dans quelle mesure les résultats observés dans

le travail qualitatif peuvent être généralisés à une plus grande population en utilisant un échantillon statistiquement représentatif et, l'analyse statistique peut être employée pour contrôler les caractéristiques du ménage et les conditions socio-économiques de différents secteurs d'étude éliminant ainsi des explications alternatives des résultats observés.

Il y a plusieurs avantages à utiliser des approches intégrées dans la recherche faite par Bamberger (2000) qui s'applique aussi aux évaluations d'impact. Parmi ceux-ci :

- Les contrôles de cohérence peuvent être intégrés à l'aide des procédures de triangulation qui permettent à deux ou plusieurs évaluations indépendantes d'être faites pour des variables clés (comme le revenu, les avis sur des projets, les raisons d'employer ou non des services publics et l'impact spécifique d'un projet).
- Des perspectives différentes peuvent être obtenues. Par exemple, bien que les chercheurs puissent considérer le revenu ou la consommation comme indicateurs clés de bien-être du ménage, les études de cas peuvent révéler que les femmes sont plus concernées par la vulnérabilité (définie comme le manque d'accès aux systèmes d'appui social dans les temps de crise, l'absence de pouvoir ou l'exposition à la violence).
- L'analyse peut être conduite à des niveaux différents. Les méthodes d'enquête peuvent fournir les bonnes évaluations du bien-être de l'individu, du ménage et au niveau de la communauté, mais ils sont beaucoup moins efficaces pour analyser des processus sociaux (conflit social, raisons d'employer ou de ne pas employer des services, etc.) ou pour l'analyse institutionnelle (comment fonctionnent efficacement la santé, l'éducation, le crédit et d'autres services et comment ils sont perçus par la communauté). Il y a beaucoup de méthodes qualitatives conçues pour analyser des questions comme le processus social, le comportement institutionnel, la structure sociale et le conflit.
- Des possibilités peuvent être offertes à la rétroaction pour aider à interpréter les résultats. Les rapports d'enquête comportent fréquemment des références aux incohérences apparentes dans les résultats ou aux différences intéressantes entre communautés ou groupes que l'on ne peut pas expliquer par des données. Dans la plupart des recherches quantitatives, une fois la phase de collecte de données achevée, il n'est pas possible de retourner sur le terrain pour vérifier de telles questions. La plus grande flexibilité de la recherche qualitative réside dans la possibilité de retourner sur le terrain pour recueillir des données complémentaires. Des enquêteurs emploient aussi des méthodes qualitatives pour vérifier des points extrêmes - réponses qui divergent des modèles généraux. Dans de nombreux cas, l'analyste de données doit prendre une décision arbitraire comme pour savoir si un ménage ou une communauté qui présente des conditions qui sont significativement au-dessus ou au-dessous de la norme doivent être exclus (en supposant qu'ils reflètent une erreur de rapport) ou si les chiffres doivent être ajustés. Les méthodes qualitatives permettent un suivi rapide sur le terrain pour vérifier ces cas.

En pratique la combinaison des méthodes quantitatives et qualitatives devrait être effectuée à chaque étape de l'évaluation. Le chapitre 2 mentionne plusieurs possibilités pour le faire. Par exemple la Réforme de l'Autonomie de l'Ecole au Nicaragua fournit un bon exemple de méthodes intégrées. Des méthodes quantitatives après une conception quasi expérimentale ont été employées pour déterminer le rapport entre la gestion décentralisée et l'étude et généraliser des résultats pour les différents types d'école. En plus des techniques qualitatives comprenant une série d'interviews clés d'information et des discussions du groupe cible avec le personnel local de différentes écoles et des parents d'élèves ont été utilisées pour analyser le contexte dans lequel la Réforme a été présentée, examiner la dynamique du processus de décision dans chaque école et évaluer les perspectives de différents acteurs de la communauté scolaire quant au processus d'autonomie (voir Annexe 1.11).

Autres Approches pour l'Evaluation de l'Impact

Deux autres termes sont particulièrement appropriés pour la discussion de l'évaluation de l'impact des projets sur la pauvreté : (a) des approches de la mesure de l'impact des Programmes d'Ajustement Structurel et (b) des évaluations à base de théorie. Tous les deux intègrent beaucoup de méthodologies traitées ci-dessus, mais chacun emploie une approche différente.

Evaluation des Programmes d'Ajustement Structurel. Il y a eu un débat substantiel sur l'impact des programmes d'ajustement structurel sur le pauvre. Une grande partie des justifications utilisées pour soutenir ce débat est, cependant, basée sur des méthodes et des hypothèses déficientes. Tout comme d'autres projets, les changements de politique au titre des Programmes d'Ajustement Structurel doivent être (a) comparés avec des hypothèses contre factuelles adaptées aux mêmes contraintes macro-économiques et (b) analysés dans le contexte de la structure économique locale et basés sur l'information empirique tirée d'enquêtes sur le ménage. Ceci, cependant est très difficile pour trois raisons. D'abord, les changements de politique peuvent avoir un impact économique de grande envergure ne permettant pas de trouver les groupes de comparaison inchangés. Deuxièmement, à cause des facteurs exogènes, des retards, des rétroactions et des substitutions, n'importe quels changements du bien-être des pauvres doivent être interprétés avec une précaution extrême. Et troisièmement, il est difficile de prévoir ce qui serait arrivé si l'ajustement n'avait pas eu lieu - quelles politiques alternatives un gouvernement aurait pu poursuivre et quel en serait l'impact sur le pauvre.

Dans la littérature, plusieurs approches ont été utilisées, chacune avec ses propres défauts. Les techniques sont dans de nombreux cas semblables à celles décrites dans l'encadré 1.2, bien que, comme indiqué dans l'encadré 1.3, évaluer la comparaison contre factuelle nécessite de grandes hypothèses qui peuvent considérablement affecter la validité des résultats. Ceci est traité de la façon la plus viable en isolant les changements de politiques spécifiques

qui affecteraient la population, comme la politique de taux de change, la politique commerciale, des réductions de dépenses publiques et des réductions de l'emploi du secteur public. Cependant, même avec cette approche il peut être difficile d'isoler l'impact de politiques spécifiques. En guise d'exemples, voir Killick (1995), Poppele, Summarto et Pritchett (1999), Bourguignon, de Melo et Suwa (1991) et Sahn, Dorosh et Younger (1996).

Encadré 1.3 Résumé des Méthodes Employées pour Evaluer des Politiques d'Ajustement Structurel

Approches sans comparaison contre factuelle

- Des études qualitatives qui évaluent les conditions de la population (identification des sous-groupes vulnérables) avant, pendant, après que des politiques d'ajustement soient mises en oeuvre à travers des groupes cibles, des interviews et d'autres techniques qualitatives.
- "Avant et Après", qui compare la performance des variables clés pendant et après un programme avec celles avant le programme. L'approche emploie des méthodes statistiques pour évaluer s'il y a un changement significatif de quelques variables essentielles avec le temps. Cette approche engendre souvent des résultats biaisés parce qu'elle suppose que, si cela n'avait pas été le programme, les indicateurs de performance auraient pris leurs valeurs de période d'avant crise.

Approches qui Génèrent une Comparaison Contre Factuelle Utilisant des Hypothèses Multiples

- Des modèles d'équilibre généraux calculables (CGEs) qui tentent d'opposer les résultats dans les groupes de traitement et de comparaison par des simulations. Ces modèles cherchent à tracer l'opération de l'économie réelle et sont généralement basés sur des matrices de comptabilité sociales détaillée rassemblées à partir des données sur des comptes nationaux, des enquêtes de dépense, du ménage et d'autres données d'enquête. Des modèles CGE génèrent souvent des résultats pour la comparaison contre factuelle bien que la force du modèle soit entièrement tributaire de la validité des hypothèses. Cela peut être problématique étant donné que des bases de données sont souvent incomplètes et la plupart des paramètres n'ont pas été estimés par des méthodes économétriques formelles. Des modèles CGE prennent aussi beaucoup de temps, sont encombrants et reviennent chers à générer.

(L'encadré continue sur la page suivante)

Encadré 1.3 (suite)

- Avec et sans des comparaisons qui analysent le comportement dans des variables clés dans un échantillon de pays de programme avec leur comportement dans des pays sans programme (un groupe de comparaison). C'est une approche à la question contre factuelle exploitant les expériences du groupe de comparaison comme mandataire pour ce qui serait autrement arrivé dans les pays du programme. Il est, cependant, tout à fait difficile de réaliser un vrai groupe de comparaison. La méthode suppose que seule l'adoption d'un programme distingue un pays de programme du groupe de comparaison et, que, l'environnement externe affecte les deux de la même manière.
- Des vérifications statistiques consistant en des régressions qui contrôlent les différences dans les conditions et les politiques entreprises dans les pays avec ou sans programme. L'approche identifie les différences entre le programme et des pays sans programme dans la période de pré-programme et contrôle en suite ces différences statistiquement pour identifier les impacts isolés des programmes dans la performance de post-réforme.

L'évaluation à base de théorie. La prémisse des évaluations à base de théorie est que les programmes et projets sont basés sur une théorie explicite ou implicite sur comment et pourquoi un programme va réussir. L'évaluation serait alors basée sur l'évaluation de chaque théorie et hypothèse d'un programme pendant la mise en œuvre plutôt qu'à mi-parcours ou après que le projet a été achevé. Dans la conception de l'évaluation, la théorie sous-jacente est présentée en plusieurs micro étapes avec les méthodes ensuite élaborées pour la collecte des données et l'analyse pour le déroulement des hypothèses. Si les événements ne se passent pas comme prévu l'évaluation peut indiquer avec une certaine confiance où, pourquoi et comment la défaillance est arrivée.

L'approche met l'accent sur les réponses des participants aux activités du programme. Les théories dirigent l'attention de l'évaluateur sur des types d'impacts probables à court et à long terme. Parmi les avantages, il y a, d'abord que l'évaluation fournit des indications rapides de l'efficacité du programme pendant la mise en œuvre du projet. S'il y a des défaillances pendant la mise en œuvre, il est possible de les situer le long du parcours. Deuxièmement l'approche aide à expliquer comment et pourquoi les effets se sont produits. Si les événements se passent comme prévu l'évaluation peut indiquer avec une certaine confiance comment les effets se sont produits. En suivant l'ordre des étapes, il est possible de suivre à la trace les micro-étapes qui ont conduit des ressources du programme jusqu'aux résultats.

Les défauts de l'approche sont semblables à beaucoup d'autres méthodologies. Particulièrement : (a) des hypothèses identifiables et des théories peuvent être en soi complexes, (b) les évaluateurs peuvent rencontrer des problèmes pour mesurer chaque étape, à moins que les données et les instruments corrects ne soient disponibles, (c) des problèmes peuvent surgir dans l'appréciation de l'effort parce que les déclarations théoriques peuvent être trop générales pour permettre une appréciation nette, et (d) il peut y avoir les problèmes d'interprétation qui rendent difficile la généralisation des résultats (voir Weiss).

Un exemple de la technique à base de théorie est piloté par le Département des Opérations et d'Évaluation de la Banque Mondiale pour évaluer l'impact des fonds d'investissement social sur les processus de prise de décision au niveau de la communauté, les structures de pouvoir traditionnelles et les relations et la capacité de la communauté, la confiance et le bien-être. Ce sera basé sur la théorie que les groupes prioritaires peuvent efficacement mettre en œuvre un projet, le faire fonctionner et sauvegarder l'investissement créé par le projet. Un jeu des hypothèses principales et des sous-hypothèses a été exposé et sera évalué en employant des données existantes d'enquêtes sur les ménages aussi bien qu'un instrument d'enquête particulièrement conçu pour un échantillon plus restreint et des groupes cibles et d'autres techniques PRA. L'information de chacune de ces sources de données sera exploitée triangulairement.

Analyse du coût bénéfique ou de rentabilité

Tandis que ce type d'analyse n'est pas strictement concerné par des mesures d'impact, il permet aux décideurs de mesurer l'efficacité d'un programme en comparant des interventions alternatives sur la base du coût de production d'un résultat donné. Il peut énormément accroître les implications de la politique de l'évaluation de l'impact et doit donc aussi être inclus dans la conception de n'importe quelle évaluation d'impact (Pour une discussion plus complète sur l'analyse du coût bénéfique et de rentabilité, voir le *Manuel sur l'Analyse Économique des Opérations et d'Investissement* de la Banque Mondiale 1996.)

L'analyse du coût bénéfique essaie de mesurer l'efficacité économique des dépenses d'un programme par rapport aux bénéfices du programme, en termes monétaires. Pour beaucoup de projets, particulièrement, dans les secteurs sociaux, il n'est pas possible de mesurer tous les bénéfices en termes monétaires. Par exemple, les bénéfices d'un programme de fournitures scolaires (livres, matériels didactiques) devraient être une éducation accrue. Au lieu de mesurer les résultats monétaires, les résultats de réalisation de l'éducation pourraient être exploités pour évaluer quantitativement les bénéfices, ce qui exigerait l'analyse de la rentabilité. Les concepts pour les deux types d'analyse sont les mêmes.

Les principales étapes de l'analyse du coût bénéfique ou de rentabilité doivent identifier toutes les dépenses et tous les bénéfices d'un projet et calculer ensuite une proportion du coût par rapport à l'efficacité. Dans le calcul des dépenses, la valeur de l'intervention

doit être incluse aussi bien que toutes les autres dépenses, comme l'administration, la prestation de services, les dépenses d'investissement (escomptées à la valeur présente nette), la valeur monétaire des marchandises ou services gratuitement fournis, les dépenses sociales comme la détérioration environnementale et les risques pour la santé. Les bénéfices peuvent être monétaires comme le gain ou le nombre d'unités livrées, des résultats de test ou des améliorations de santé. Quand les bénéfices ne peuvent être évalués quantitativement, il est possible d'employer des indicateurs subjectifs comme les systèmes d'indemnité ou de classement. Cette approche cependant peut être biaisée dans l'interprétation des résultats subjectifs.

Une fois que les dépenses et bénéfices ont été arrêtés, le ratio de rentabilité (R) est donc $R = \text{coût/unité (ou bénéfice)}$. Ce ratio peut ensuite être comparé à travers des interventions pour mesurer l'efficacité. En théorie, cette technique est tout à fait directe. En pratique, cependant, beaucoup de risques impliqués dans l'identification et la quantification des coûts et bénéfices. Il est important de s'assurer que des indicateurs appropriés soient choisis, que les méthodologies et les hypothèses économiques employées soient cohérentes à travers les ratios, et que les ratios soient effectivement comparables. Et comme dans d'autres techniques employées dans l'analyse de l'impact, l'évaluation de la rentabilité peut être mieux faite quand on l'inclut dans la conception de l'évaluation dès les toutes premières étapes. Cela permet de collecter l'information sur le coût et bénéfice nécessaire et d'assurer la cohérence.

Le choix d'une méthodologie

Compte tenu la diversité des types de projet, des questions d'évaluation, de la disponibilité des données, le coût, les contraintes de temps et les situations de pays, chaque étude d'évaluation d'impact sera différente et nécessitera une certaine combinaison de méthodologies appropriées, à la fois quantitatives et qualitatives. L'évaluateur doit soigneusement explorer les options méthodologiques dans la conception de l'étude dans le but d'obtenir les résultats les plus fiables possibles. Parmi les méthodes quantitatives on considère les conceptions expérimentales et les comparaisons harmonisées comme la seconde meilleure alternative. D'autres techniques, cependant, peuvent aussi produire des résultats fiables, particulièrement avec une bonne conception de l'évaluation et des données de haute qualité.

S'il est évident que les évaluations "Meilleure Pratique" passées en revue pour ce manuel fait ressortir que le choix des méthodologies d'évaluation d'impact n'est pas mutuellement exclusif. En effet des évaluations plus fiables combinent souvent des méthodes afin d'assurer la fiabilité et prévoir des éventualités dans la mise en œuvre. La combinaison des approches « avec et sans » et « avant et après » utilisant la fusion des données de base et des données de suivi constitue une combinaison fortement recommandée du point de vue méthodologique (Subbarao, et autres 1999). La disponibilité des données de base permettra aux évaluateurs de vérifier l'intégrité des groupes de traitement et des groupes de comparaison, d'évaluer le ciblage et de se préparer pour une évaluation d'impact fiable. Ceci est vrai même pour

des conceptions de simulation aléatoire. Bien que la simulation aléatoire assure des groupes cibles et de comparaison équivalente au moment du choix, cette particularité ne doit pas pousser les évaluateurs à penser qu'ils n'ont pas besoin de données de base. En effet, les données de base peuvent être cruciales pour rétablir pourquoi certains événements ont eu lieu et faire un contrôle pour ces événements dans l'évaluation d'impact.

Il est aussi fortement recommandé que le coût-bénéfice ou la rentabilité soit incorporé dans l'analyse. Cette méthodologie peut permettre aux décideurs de comparer des interventions alternatives sur la base du coût de production d'un résultat donné. Ceci est particulièrement important dans le contexte des pays en développement où les ressources sont extrêmement limitées.

Finalement, la combinaison des méthodes quantitatives et qualitatives est l'idéal parce qu'elle fournira l'impact quantifiable d'un projet aussi bien qu'une explication des processus et des interventions qui ont généré ces résultats. Bien que chaque évaluation d'impact ait des caractéristiques uniques nécessitant des approches méthodologiques différentes, certaines caractéristiques considérées comme bonnes pratiques comprennent :

- une estimation des comparaisons contre factuelles faite par (a) l'utilisation de la nomination aléatoire pour créer un groupe témoin (la conception expérimentale) et (b) l'utilisation appropriée et soignée d'autres méthodes comme la mise en parallèle pour créer un groupe témoin (la comparaison quasi expérimentale)
- le contrôle des différences d'avant et d'après programme chez les participants et l'observation des signes des impacts de programme ; s'assurer qu'il y a des données appropriées collectées au départ et pendant le suivi (incluant des délais suffisants pour permettre de tenir compte des impacts d'un programme).
- Les groupes cibles et les groupes de contrôle sont de tailles suffisantes pour établir des inductions statistiques avec une usure minimale.
L'analyse coût-bénéfice ou de rentabilité est réalisée pour mesurer l'efficacité du projet.
- Des techniques qualitatives sont intégrées pour permettre la triangulation des résultats.

Chapitre 2

Démarches Clés dans la Conception et la Mise en œuvre des Evaluations d'Impact

Entreprendre une étude d'évaluation d'impact peut être stimulant et coûteux avec des questions relatives à la mise en œuvre qui surgissent continuellement. Ces défis mettent en évidence l'importance d'une étude bien conçue, une équipe engagée et hautement qualifiée et une bonne communication entre les membres de l'équipe d'évaluation. En incorporant dès le départ l'évaluation dans la conception d'un projet, il sera possible d'obtenir des résultats en temps opportun pour que les conclusions puissent être utilisées pour les réajustements à mi-parcours des composantes spécifiques.

Indépendamment de la taille, du type de programme ou de la méthodologie employée pour l'évaluation, il y a plusieurs démarches clés à effectuer comme décrit ci-dessous (Encadré 2.1) Ce chapitre fournira un exposé de ces démarches ainsi qu'un exposé des nombreuses questions qui peuvent surgir dans la mise en œuvre. La succession de ces démarches est critique particulièrement pour assurer la collecte des données nécessaires avant que le projet n'entre en exécution. La planification faite à temps fournit l'occasion de faire une simulation aléatoire, de construire a priori des comparaisons parallèles, de collecter des données de base et d'identifier des enquêtes qui pourraient être utilisées dans une approche associant des résultats de tendance.

Tout le travail de conception et de collecte des données initiales doit se faire pendant l'identification et la préparation du projet. Idéalement, quelques résultats seront disponibles pendant la mise en œuvre du projet de façon qu'ils puissent servir à l'amélioration de la conception du projet si nécessaire. Un bon exemple de plans d'évaluation incorporés dès les toutes premières étapes est illustré dans le Projet de Nutrition et de Développement de la Petite Enfance en Ouganda.

Déterminer l'opportunité d'effectuer une évaluation

Une première décision est de savoir si une évaluation d'impact est nécessaire ou non. Comme discuté ci-dessus les évaluations d'impact diffèrent des autres évaluations en ce qu'elles sont centrées sur l'évaluation de la causalité. Etant donné la complexité et le coût de l'exécution de l'évaluation d'impact, les coûts et les bénéfices doivent être évalués et l'utilisation d'une autre approche qui serait plus appropriée, comme le contrôle des indicateurs clés de performance ou une évaluation du processus, devrait également être envisagée. Ces approches ne devraient pas

* Ce chapitre s'est beaucoup inspiré d'un exposé préparé par Laura Rawlings, *Implementation Issues in Impact Evaluation*, Processed, 1999.

Encadré 2.1 Principales démarches dans la conception et l'exécution des évaluations d'impact

Pendant l'Identification et la Préparation du Projet

1. Déterminer s'il faut nécessairement effectuer une évaluation ou non
2. Clarifier les objectifs de l'évaluation
3. Explorer la disponibilité des données
4. Concevoir l'évaluation
5. Former l'équipe d'évaluation
6. Si des données seront collectées :
 - (a) Conception et sélection des échantillons
 - (b) Développement de l'instrument de collecte des données
 - (c) Recruter et former le personnel de terrain
 - (d) Réaliser l'enquête pilote
 - (e) Collecte de données
 - (f) Gestion et accès aux données

Pendant la Supervision du Projet

7. Poursuivre la collecte des données
8. Analyser les données
9. Rendre compte des résultats par écrit et les discuter avec les décideurs et les bailleurs de fonds
10. Incorporer les résultats à la conception du projet.

être considérées comme des substituts aux évaluations d'impact; en fait, elles constituent souvent des compléments critiques aux évaluations d'impact. Et peut-être les raisons décisives d'effectuer une évaluation ou non dépendent de l'importance de l'appui politique et financier.

L'effort complémentaire et les ressources nécessaires pour conduire des évaluations d'impact sont mieux mobilisés quand le projet est innovateur, transposable et implique des allocations de ressources substantielles et comporte des interventions bien définies. Par exemple, l'évaluation de l'impact du Fonds d'investissement social bolivien a rempli chacun de ces critères. D'abord, on a considéré comme innovateur et transposable le nouveau modèle de fonds social présenté en Bolivie, deuxièmement le fonds social a été responsable d'à peu près 25 pour cent de tous les investissements publics en Bolivie depuis le début de l'évaluation et troisièmement les interventions étaient bien définies par le menu de micro-projets du fonds social.

On doit aussi donner la priorité aux évaluations d'impact si le projet en question lance une nouvelle approche comme un programme pilote qui sera plus tard pris en considération pour une extension basée sur les résultats d'évaluation ou les nouveaux Prêts pour l'Apprentissage et l'Innovation de la Banque Mondiale. Ce raisonnement a fait de la réforme de l'autonomie scolaire nicaraguayenne un bon candidat à une évaluation d'impact. L'étude d'évaluation a accompagné l'expérimentation par le gouvernement d'un nouveau modèle de gestion scolaire décentralisée depuis sa phase pilote au milieu des années 1990 jusqu'à son extension à presque tous les collèges d'enseignement général et environ la moitié de toutes les écoles primaires aujourd'hui. L'évaluation a été gérée par une équipe internationale en étroite collaboration comprenant le personnel local de l'unité de recherche et d'évaluation du Ministère de l'Education Nationale et l'unité d'évaluation et le bureau de coordination du Projet d'Education Primaire de la Banque Mondiale à Managua. Leur implication a assuré que l'évaluation a informé les décisions de politique clés relatives à l'évaluation quant à la modification et l'extension du projet pilote.

Une autre considération importante est de s'assurer que le programme qui doit être évalué est suffisamment développé pour être soumis à une évaluation d'impact. Les projets pilotes et les réformes naissantes sont souvent enclins aux révisions quant à leur contenu aussi bien que pour savoir comment, quand et par qui, ils seront mis en œuvre. Ces changements peuvent saper la cohérence de l'effort d'évaluation, des conceptions particulièrement expérimentales et d'autres types d'évaluation prospective qui reposent sur les données de base et de suivi des groupes de traitement et de contrôle clairement établis. Là où la politique à évaluer est encore en cours de définition, il peut être recommandé d'éviter d'utiliser une évaluation d'impact afin de permettre la flexibilité dans le projet

L'obtention de l'appui des décideurs et des financiers pour une évaluation d'impact peut être stimulante mais est un préalable nécessaire pour avancer. Ils doivent être convaincus que l'évaluation est un exercice utile traitant des questions qui seront pertinentes aux décisions concernant l'affinement, l'extension ou la réduction du programme évalué. Ils doivent aussi être convaincus de la légitimité de la conception de l'évaluation et donc des résultats, particulièrement quand ceux-ci ne sont pas aussi positifs que prévu.

Le financement d'une évaluation d'impact reste une question difficile pour les directeurs de programme, de même que pour les partenaires bénéficiaires. La question du financement se compose du fait que les données sur des dépenses d'évaluation sont d'habitude difficile à obtenir. Et peut-être la question la plus difficile vient de la bonne valeur publique de l'évaluation : si les résultats de l'évaluation vont être employés pour informer la politique appliquée à l'extérieur des frontières nationales où l'évaluation est conduite, comme c'est souvent le cas, pourquoi un seul pays devrait-il supporter le coût de l'évaluation? Parmi les études de cas qui avaient l'information sur les sources de financement, l'information montre que les pays assument souvent la majorité, mais pas l'intégralité, des dépenses d'évaluation. Comme discuté plus pleinement dans le chapitre 4, beaucoup de cas passés en revue suggèrent que l'exécution réussie d'une évaluation d'impact exige

non seulement un engagement de ressources substantielles des pays bénéficiaires mais aussi l'implication du personnel de la Banque Mondiale ou de chercheurs et consultants externes, nécessitant des ressources au-delà de celles fournies par le pays.

Clarifier les objectifs de l'évaluation

Une fois qu'il a été décidé qu'une évaluation d'impact est appropriée et justifiée, établir des objectifs clairs qui seront le point central de l'évaluation contribuera grandement à son succès. Des objectifs clairs sont essentiels pour l'identification des besoins de l'information, pour la fixation des résultats et des indicateurs d'impact et bâtir une stratégie solide d'évaluation pour trouver des réponses aux questions posées. L'utilisation d'une approche de cadre logique fournit un bon outil généralement employé pour identifier les buts et les besoins d'information du projet autour desquels l'évaluation peut être bâtie.

Le cadre logique, de plus en plus employé à la Banque Mondiale, est basé sur une matrice simple "quatre par quatre " qui associe l'information avec des objectifs du projet et la façon dont l'exécution sera suivie à la trace, utilisant des jalons et des programmes de travail, quel impact les résultats du projet auront sur l'institution ou le système bénéficiaire et comment cela sera mesuré et comment les ressources sont utilisées pour produire des résultats (voir Annexe 5 pour des exemples). Autrement dit, il est assumé que l'impact du projet prévu est fonction des résultats du projet aussi bien qu'une série d'autres facteurs. Les résultats sont une fonction des entrées du projet qui sont à leur tour fonction des ressources du projet et de facteurs extérieurs au projet. Des mesures quantifiables doivent alors être identifiées pour chaque maillon du cycle du projet. Cette approche n'interdit pas à l'évaluateur de considérer aussi les impacts fortuits d'un projet mais sert à garder les objectifs clairs et centrés. Des techniques qualitatives sont aussi utiles dans la révélation de la participation dans la clarification des objectifs de l'évaluation et des indicateurs d'impact y afférents.

Bien qu'une déclaration d'objectif puisse sembler, en apparence, une des parties les plus faciles du processus d'évaluation, sera peut être extrêmement difficile. Par exemple des déclarations qui sont trop vagues ne se prêtent pas à l'évaluation. La déclaration d'objectif dans l'évaluation du PROBECAT au Mexique (Annexe 1.9) selon laquelle l'évaluation concerne l'effet du programme de formation PROBECAT sur les résultats du marché du travail serait plus précise si elle avait été ramenée à l'effet de PROBECAT sur les heures de travail, les indemnités horaires, le salaire mensuel et la période du premier placement au travail pour les différents types d'ouvriers. L'évaluation du PROGRESA au Mexique fournit un bon exemple et crée un profit clair des objectifs multiples dès le début avec une discussion séparée de chaque composante avec des objectifs détaillés dans les sous catégories

(Annexe 1.10). Il est particulièrement important parce que

l'intervention était tout à fait complexe avec une évaluation devant traiter non seulement l'impact du programme mais aussi les aspects de ciblage et du chronogramme des opérations.

L'examen d'autres composantes de l'évaluation telles que le coût rentabilité ou le processus d'évaluations peut aussi constituer des objectifs importants d'une étude et peut compléter l'évaluation d'impact. La rentabilité peut être d'un intérêt particulier pour des décideurs dont il s'agira de réduire, étendre ou reformer l'intervention évaluée. Sur des questions liées à la prestation de service, une évaluation du processus pourrait être appropriée pour les procédures, la dynamique, les normes et les contraintes dans lesquelles un programme particulier est réalisé.

Explorer la disponibilité des données

Plusieurs types de données peuvent être utilisés pour effectuer des études d'évaluation d'impact. Ceux-ci peuvent comprendre une gamme d'enquêtes allant des enquêtes de sessions croisées ou d'ensemble, aux interviews qualitatives à finalité ouverte. Idéalement, cette information est disponible au niveau individuel pour assurer que le vrai impact peut être évalué. L'information au niveau des ménages peut cacher l'allocation de ressources internes au ménage qui affecte des femmes et des enfants parce qu'ils ont souvent un accès plus limité aux ressources productives du ménage. Dans de nombreux cas, l'évaluation d'impact profitera de quelques données existantes ou se superposera à une enquête en cours, ce qui peut considérablement réduire les coûts. Avec cette approche, cependant, les problèmes peuvent surgir dans le choix du temps de collecte des données et avec la flexibilité de la conception du questionnaire. L'Encadré 2.2 met en lumière quelques points clés à retenir dans l'exploration de l'utilisation des ressources des données existantes pour l'évaluation d'impact.

Avec cette créativité, il peut être possible de maximiser des ressources d'information existantes. L'évaluation du Fonds d'Investissement social hondurien (voir le chapitre 4) est un bon exemple. Cette étude a utilisé un module tiré de l'enquête nationale sur les revenus dans le questionnaire du fonds social, permettant ainsi au revenu des bénéficiaires du fonds social d'être comparé avec des mesures nationales pour évaluer le ciblage de la pauvreté (Walker et d'autres 1999).

Au plus bas niveau, des données sur l'univers de la population concernée seront exigées comme base à partir de laquelle déterminer les tailles types, construire le cadre de l'échantillonnage et choisir l'échantillon. D'autres types de données qui peuvent être disponibles dans un pays donné et peuvent être employés pour des évaluations d'impact différentes comprennent (voir Valadez et Bamberger 1994) : revenu du ménage et enquêtes sur les revenus et les dépenses des ménages, étude de la mesure du niveau de vie, enquêtes sur le marché du travail ; rapports des coopératives, associations de coopérative de crédit et d'autres institutions financières ou rapport scolaire sur l'état de la fréquentation, du redoublement et des examens; rapports de santé publique sur la mortalité infantile,

Encadré 2.2 Points clés pour identifier des ressources de données pour l'évaluation d'impact

- Connaissez bien le programme. Il est risqué de s'engager dans une évaluation sans bien connaître les détails administratifs et institutionnels du programme ; cette information vient typiquement de l'administration du programme.
- Collectez l'information sur les faits stylisés appropriés par rapport à la détermination. Les faits appropriés pourraient inclure la carte de la pauvreté, comment fonctionne le marché du travail, les divisions ethniques majeures et d'autres programmes publics appropriés.
- Soyez éclectique au sujet des données. Les sources peuvent embrasser tant des interviews informelles, peu structurées avec des participants au programme que des données quantitatives d'échantillons représentatifs. Cependant, il est extrêmement difficile de poser des questions contre factuelles dans des interviews ou à des groupes ciblés ; essayer de demander à quelqu'un qui participe actuellement à un programme public : " que feriez-vous maintenant si ce programme n'avait pas existé ? " Parler aux participants au programme peut être, mais il est improbable d'obtenir une évaluation crédible sur cette seule base.
- Assurez-vous qu'il y a des données sur les indicateurs de résultat et des variables explicatives appropriées. Ces dernières ont besoin de traiter l'hétérogénéité dans les hypothèses de résultat sur la participation au programme. Les résultats peuvent différer en fonction, par exemple, de l'instruction. Il n'est pas possible de voir l'impact du programme à moins qu'on contrôle cette hétérogénéité.
- Selon les méthodes utilisées, les données pourraient aussi être nécessaires sur des variables qui influencent la participation, mais n'influencent pas la participation donnée aux résultats.
- Ces variables instrumentales peuvent être valables pour trier les effets suscités possibles de programmes non aléatoires (encadré 1.2).
- Les données sur les résultats et d'autres variables explicatives appropriées peuvent être quantitatives ou qualitatives. Mais il doit être possible d'organiser l'information dans toute sorte de structure de données systématiques. Un exemple simple et commun est que l'on a les valeurs de variables diverses incluant un ou plusieurs indicateurs de résultat pour des unités d'observation diverses (des individus, des ménages, des sociétés, des communautés)
- Les variables sur lesquelles on a des données et les unités d'observation qu'on utilise sont souvent choisies comme une partie de la méthode d'évaluation. Ces choix

(L'encadré continue sur la page suivante).

Encadré 2.2 (*Suite*)

devraient être ancrés à la connaissance antérieure du programme (ses objectifs, bien sûr, mais aussi comment il est dirigé) et l'arrangement dans lequel il est présenté.

- La source spécifique des données sur les résultats et leurs déterminants, y compris la participation au programme, vient typiquement des données d'enquête de toute sorte. L'unité d'observation pourrait être le ménage, la société ou le secteur géographique selon le type de programme que l'on étudie.
- Les données d'enquête peuvent souvent être complétées avec d'autres données utiles sur le programme (telle la donnée de base de contrôle du programme) ou l'arrangement (telles les données de base géographiques).

Incidence des différentes maladies infectieuses, le nombre de femmes sollicitant conseil sur la contraception, l'utilisation du préservatif ; des enquêtes spécialisées menées par des universités, des organisations non gouvernementales et des groupes consultatifs, contrôle de données des administrateurs de programme et étude de cas de projet.

Exploitation des données d'enquêtes existantes. Plusieurs enquêtes peuvent être en cours de planification ou en cours d'exécution. Si une enquête évaluant les indicateurs nécessaires est projetée, l'évaluation peut être en mesure de prendre l'échantillonnage de la population concernée au cours de l'enquête générale (par exemple, destiné à être utilisé dans l'approche des résultats de tendance) comme ce fut le cas pour l'évaluation du Fonds d'investissement social nicaraguayen et l'évaluation du programme protection-travail TRABAJAR en Argentine (Jalan et Ravallion 1998). Par contre, si une enquête est projetée qui couvrira la population concernée, l'évaluation peut être en mesure d'introduire une question ou la série de questions comme une partie de l'enquête ou ajouter une enquête qualitative pour compléter l'information quantitative. Par exemple, le Crédit pour le programme d'éducation au Ghana a inclus un jeu d'interviews qualitatives avec des partenaires clés aussi bien qu'avec des groupes cibles non participants et participants qui ont fourni la confirmation qualitative des résultats quantitatifs (Annexe 6). L'évaluation a mesuré l'impact du programme sur le statut institutionnel et la sécurité alimentaire des ménages pauvres. Des données qualitatives ont inclus des questions spécifiques sur le revenu et les dépenses du ménage, et le niveau de compétence tandis que des données qualitatives se sont concentrées sur le renforcement des capacités des femmes - statut et prise de décision dans les ménages, réseaux sociaux, confiance en soi et ainsi de suite.

Conception de l'évaluation

Une fois que les objectifs et les ressources de données sont clairs, il est possible de commencer la phase de conception de l'étude de l'évaluation d'impact. Le choix des méthodologies dépendra des questions de l'évaluation, du choix du temps de l'évaluation, des contraintes du budget et de la capacité de mise en œuvre. Les avantages et les inconvénients des différents types de conception traités au chapitre 1 doivent être équilibrés pour déterminer les méthodologies les plus appropriées et la manière dont des techniques quantitatives et qualitatives peuvent être intégrées pour se compléter les unes les autres.

Même après que la conception de l'évaluation a été décidée et intégrée au projet, les évaluateurs doivent être préparés à être flexibles et faire des modifications à la conception pendant que le projet est mis en œuvre. En plus des dispositions doivent être prises pour suivre à la trace les interventions du projet si l'évaluation inclut des données de base et de suivi pour que le travail d'évaluation soit parallèle à l'allure réelle du projet.

Dans la définition de la conception, il est aussi important de déterminer comment l'évaluation d'impact s'ajustera à la plus large stratégie de contrôle et d'évaluation appliquée au projet. Tous les projets doivent être contrôlés pour que les administrateurs, les bailleurs de fonds et les décideurs puissent suivre l'évolution du programme comme il se déroule. L'effort de l'évaluation, comme traité ci-dessus, doit être façonné aux exigences d'information du projet.

Question de l'évaluation. Les questions d'évaluation posées sont étroitement liées à la conception de l'évaluation en termes de type de données collectées, unité d'analyse, méthodologies employées et temps choisi pour les diverses étapes. Par exemple, dans l'évaluation de l'impact des manuels scolaires sur les résultats de l'apprentissage, il serait nécessaire de façonner l'évaluation pour mesurer l'impact sur les étudiants, les salles de classe et les enseignants pendant une année scolaire donnée ; ce qui serait très différent de la mesure de l'impact des services fournis par des investissements de fonds sociaux qui exigeraient des données sur les équipements de la communauté et sur les ménages. Les études de cas dans l'Annexe 1 fournissent les autres exemples en la manière dont la question de l'évaluation peut affecter la conception de l'évaluation.

Dans la clarification des questions d'évaluation, il est aussi important de considérer les implications de genre sur l'impact du projet. Au début, cela ne peut pas toujours être évident, cependant, dans la mise en œuvre du projet, il peut y avoir des effets secondaires sur les ménages, ce qui ne serait pas nécessairement identifié sans la collecte de données spécifiques et les efforts d'analyse.

Questions du choix du temps de l'évaluation et du budget. La question du choix du temps la plus critique est de savoir s'il est possible de commencer la conception de l'évaluation avant que le projet ne soit mis en œuvre et quand les résultats seront disponibles. Il est aussi utile d'identifier dès le départ à quels points du cycle du projet l'information

fournie par le travail de l'évaluation sera nécessaire pour que la collecte des données et les activités d'analyse puissent être liées. Avoir des résultats au bon moment peut être crucial pour les décisions - par exemple au cours d'un examen de projet, pendant une période électorale ou quand des décisions relatives à la continuation d'un projet sont en train d'être prises.

Quelques méthodes nécessitent plus de temps pour la mise en œuvre que d'autres. Le choix aléatoire des méthodes "avant et après" (par exemple des comparaisons réflexives) prend plus de temps à mettre en œuvre que des approches de mise en correspondance de la comparaison ex-post. En exploitant des approches "avant et après" qui utilisent des évaluations de base et de suivi, le temps doit être laissé au dernier membre du groupe de traitement pour recevoir l'intervention et ensuite d'habitude on laisse plus de temps pour que les effets post programme se concrétisent et soient observés. Grossman (1994) suggère que 12 à 18 mois après l'inscription de l'échantillon dans l'intervention soit une période type à observer avant l'examen des impacts. Dans des projets de la Banque Mondiale, axés sur des données de base, l'attente que l'intervention prenne place et que les résultats se concrétisent peut prendre des années. Par exemple dans l'évaluation du Fonds d'investissement social bolivien qui tablait sur les données de base collectées en 1993 des données complémentaires n'étaient collectées qu'en 1998 à cause du temps nécessité pour les interventions (projet d'adduction d'eau et de systèmes sanitaires, cliniques de santé et écoles) réalisées pour que les effets sur la santé de la population des bénéficiaires et des résultats éducatifs soient obtenus. Une période semblable de temps a été nécessaire pour l'évaluation d'un projet d'éducation primaire au Pakistan qui a utilisé une conception expérimentale avec des données d'enquêtes de base et de suivi pour évaluer l'impact des écoles communautaires sur les résultats des étudiants, y compris les performances académiques.

Les exigences du choix du temps de l'évaluation ne peuvent pas guider le projet en évaluation. Par leur nature même, les évaluations sont soumises aux délais établis par le reste du projet. Les évaluations doivent attendre pour les projets qui sont lents pour les déboursements et la réalisation des interventions. Et même si les projets avancent à l'allure établie, quelques interventions prennent plus de temps pour se réaliser comme des projets d'infrastructure. Les délais pour l'évaluation sont aussi sensibles aux indicateurs choisis parce que beaucoup d'indicateurs, comme les changements de taux de fécondité ou la réussite scolaire mettent plus de temps pour se manifester dans la population des bénéficiaires.

Capacité de mise en œuvre. Une considération finale dans l'échelle et la complexité de la conception d'évaluation est la capacité de mise en œuvre de l'équipe d'évaluation. Les questions de la mise en œuvre peuvent être très défiantes particulièrement dans les pays en voie de développement où il y a peu d'expérience de la recherche appliquée et des évaluations de programme. La composition de l'équipe d'évaluation est très importante aussi bien que l'expérience des membres de l'équipe avec les différents types de méthodologie et leur capacité quant à d'autres activités en cours d'exécution par l'unité d'évaluation. Ceci est particulièrement

vrai quand on travaille avec les agences du secteur public qui ont des responsabilités multiples et un personnel limité. La conscience que l'unité a de la charge du travail est importante pour évaluer non seulement comment elle affectera la qualité de l'évaluation menée mais aussi le coût d'opportunité de l'évaluation en ce qui concerne d'autres efforts dont l'unité est responsable. Il y a plusieurs exemples des efforts d'évaluation qui ont été dévoyés quand le personnel clé a été appelé sur d'autres projets et n'était pas ainsi en mesure de mettre en œuvre la collecte des données sur le programme au point critique dans le temps (comme un point pendant l'année scolaire ou pendant la saison agricole). De telles situations peuvent être évitées par la coordination avec des responsables dans l'unité en charge de l'évaluation pour s'assurer qu'un équilibre est réalisé en respectant le choix du temps des diverses activités aussi bien que la répartition du personnel et des ressources sur ces activités. Alternativement, il est peut-être préférable de passer un contrat avec un bureau d'études pour effectuer l'évaluation (traité ci-dessous).

Formation de l'équipe d'évaluation

Une gamme de compétences est nécessaire dans le travail d'évaluation. La qualité et l'utilité éventuelle de l'évaluation d'impact peuvent être énormément améliorées avec la coordination entre les membres d'équipe et les décideurs dès le début. Il est donc important d'identifier les membres de l'équipe le plus tôt possible, convenir des rôles et des responsabilités et établir des mécanismes pour la communication aux points clés de l'évaluation.

Dans l'équipe centrale, figurent le directeur de l'évaluation, les analystes (tant des économistes que d'autres spécialistes des sciences humaines) et pour des conceptions d'évaluation impliquant une nouvelle collecte de données, un expert de prélèvement d'échantillon, le concepteur de l'enquête, le responsable des opérations de terrain et l'équipe de terrain et des responsables et des processeurs de données (pour un guide global à la conception et l'exécution des enquêtes) voir Grosh et Muñoz 1996). Selon la taille, la portée et la conception de l'étude, certaines de ces responsabilités sont partagées ou d'autres besoins en personnel pourraient s'ajouter à cette équipe centrale. Dans les cas où les analystes de politique peuvent n'avoir pas eu l'expérience d'intégration des approches quantitatives et qualitatives, il peut être nécessaire de passer un temps complémentaire au stade initial de la formation de l'équipe pour sensibiliser les membres de l'équipe et assurer la pleine collaboration. Les responsabilités détaillées des membres de l'équipe sont les suivantes :

- Responsable de l'évaluation - Le responsable de l'évaluation est chargé de l'établissement des besoins de l'information et des indicateurs pour l'évaluation (qui sont souvent établis avec le client en employant une approche de structure logique), rédigeant les termes de référence pour l'évaluation, choisissant la méthodologie d'évaluation et identifiant l'équipe d'évaluation. Dans de nombreux cas, le responsable de l'évaluation effectuera aussi l'analyse de la politique.

- **Analyse politique** - Un économiste est nécessaire pour l'analyse quantitative aussi bien qu'un sociologue ou un anthropologue pour assurer une ressource participative et une analyse qualitative aux différentes étapes de l'évaluation d'impact. Tous les deux doivent être impliqués dans la rédaction du rapport d'évaluation.
- **Expert d'échantillonnage** - L'expert de l'échantillonnage doit être capable d'effectuer des calculs solides pour déterminer les tailles types appropriées pour les indicateurs établis, choisir l'échantillon, passer en revue les résultats de l'échantillon réel contre l'échantillon conçu et incorporer les poids de l'échantillonnage pour l'analyse. Pour des données qualitatives, l'expert du tirage des échantillons doit guider le processus de choix type en coordination avec les analystes, assurant que les procédures établies garantissent que les informateurs corrects sont choisis. L'expert du prélèvement des échantillons devrait être aussi chargé de la sélection des sites et des groupes pour l'enquête pilote et devra souvent être associée à un coordonnateur local de l'information, responsable de la collecte, pour l'expert de l'échantillonnage, des données d'où l'échantillon sera tiré.
- **Le concepteur d'enquête** - Il pourrait être une personne ou une équipe, dont la responsabilité est de concevoir les instruments de la collecte des données accompagnant des manuels et des livres codes et d'être en coordination avec le responsable d'évaluation pour assurer que les instruments de la collecte des données produiront effectivement les données nécessaires pour l'analyse. Cette personne ou équipe doit aussi être impliquée dans l'enquête pilote et l'affinement des questionnaires.
- **Le directeur des opérations de terrain et le personnel** - Le directeur doit être responsable de la supervision de tout le travail de la collecte de données, depuis la planification des voies pour la collecte des données jusqu'à la formation et l'organisation des équipes de terrain, généralement composées de superviseurs et d'intervieweurs. Les superviseurs gèrent généralement le personnel de terrain (d'habitude des intervieweurs, des opérateurs de saisie et des chauffeurs) et sont responsables de la qualité des données collectées sur le terrain. Les intervieweurs administrent les questionnaires. Dans certaines cultures, il est nécessaire d'assurer que des intervieweurs masculins et féminins effectuent les enquêtes et qu'elles soient administrées séparément par les hommes et les femmes.
- **Responsables des données et analystes** - Ces membres d'équipe conçoivent les programmes d'entrée des données, entrent les données, vérifient la validité des données, fournissent la documentation des données nécessaires et produisent les résultats de base qui peuvent être vérifiés par les analystes de données.

Dans la formation de l'équipe d'évaluation, il y a aussi certaines décisions importantes que le directeur d'évaluation doit prendre en rapport avec la capacité locale et les dispositions institutionnelles appropriées, en vue d'assurer l'impartialité et la qualité des résultats d'évaluation. La première est de savoir si la capacité locale existe pour la mise en œuvre de l'évaluation, ou de certaines parties de l'évaluation, et le genre de supervision et d'assistance extérieure qui serait nécessaire. La capacité d'évaluation varie énormément

d'un pays à un autre et bien que les contrats internationaux permettent à des sociétés d'un pays d'effectuer des évaluations dans un autre pays (un exemple est l'évaluation de PROGRESA en cours de réalisation par l'Institut International de Recherche sur l'Alimentation et la Politique), la pratique générale pour les projets soutenus par la Banque Mondiale semble être de mettre en œuvre l'évaluation associant le personnel local qui fournit une grande partie de la supervision internationale. Donc, il est nécessaire d'évaluer d'une façon critique la capacité locale et déterminer qui sera responsable de quels aspects de l'effort d'évaluation. Indépendamment de la composition finale de l'équipe, il est important de désigner un directeur d'évaluation qui sera capable de travailler efficacement avec les producteurs de données aussi bien que les analystes et des décideurs utilisant les données et les résultats de l'évaluation. Si cette personne n'est pas basée sur place, il est recommandé qu'un directeur national soit désigné pour coordonner l'effort d'évaluation en collaboration avec le directeur international.

La deuxième décision est de savoir s'il faut travailler avec une société privée ou une agence publique. Des sociétés privées peuvent être plus fiables pour ce qui est de fournir des résultats en termes de délai mais le renforcement des capacités dans le secteur public est perdu et les sociétés privées sont, de façon compréhensible, moins portées à incorporer à l'évaluation des éléments qui rendraient le travail plus coûteux. Quelle que soit la contrepartie ou combinaison de contreparties finalement réalisée, un examen approfondi des activités d'évaluation passées des collaborateurs potentiels est essentiel pour la formation d'un choix justifié.

La troisième décision est de savoir quel degré de séparation institutionnelle mettre en place entre les fournisseurs d'évaluation et les utilisateurs d'évaluation. Il y a beaucoup à gagner dans l'objectivité en faisant effectuer l'évaluation indépendamment de l'institution responsable du projet en évaluation. Cependant, les évaluations peuvent souvent avoir des buts multiples, y compris le renforcement des capacités en évaluation dans les agences du gouvernement et la sensibilisation des opérateurs de programmes aux réalités de leurs projets une fois que ceux-ci sont effectués sur le terrain. Au minimum, les utilisateurs d'évaluation qui peuvent aller des décideurs des agences du gouvernement des pays bénéficiaires aux organisations non gouvernementales, pays donateurs bilatéraux et institutions de développement internationales, doivent être suffisamment impliqués dans l'évaluation pour assurer que le processus d'évaluation soit reconnu comme légitime et que les résultats produits soient conformes à leurs besoins d'information. Autrement dit, les résultats de l'évaluation vont moins probablement être utilisés pour informer la politique. Dans l'analyse finale, le directeur d'évaluation et ses clients doivent réaliser l'équilibre juste entre l'implication des utilisateurs d'évaluations et le maintien de l'objectivité et de la légitimité des résultats.

Développement des données

Avoir des données adéquates et fiables est une ressource nécessaire à l'évaluation de l'impact d'un projet. Des données de haute qualité sont essentielles pour la validité des résultats de l'évaluation.

Comme traité ci-dessus, inventorier quelles données existent, constitue un premier pas important avant le lancement d'autres efforts de collecte de nouvelles données. Le tableau 2.1 relie les méthodologies d'évaluation de base aux besoins de données. La plupart de ces méthodologies peuvent incorporer des techniques qualitatives et participatives dans la conception de l'instrument de l'enquête, dans l'identification des indicateurs et dans la contribution à l'identification des variables de contrôle employées pour l'harmonisation, ou des variables instrumentales.

Tableau 2.1 Méthodologies d'Evaluation et Spécification des Données Correspondantes

Méthodologie	Spécification des données		Utilisation d'approche qualitative
	Minimale	Idéale	
Contrôles des données randomisées ou expérimentales	Projet unique de données comparatives avec ou sans des bénéficiaires	Enquête sur des études de bases et de suivi sur des bénéficiaires de même que des non bénéficiaires. Prévoit des rectifications d'évènements contemporains, en plus permet de surveiller les mesures de l'impact. (Permet une estimation de différence double)	<ul style="list-style-type: none"> • Informe sur la conception des instruments d'enquêtes, d'échantillonnage. • Identifie des indicateurs • Collecte et enregistre des données utilisant <ul style="list-style-type: none"> - Données textuelles - Interviews informelles et semi structurées - Réunions des communautés et des groupes cibles - Observation directe - Méthodologies participatives - Photographies - Triangulation - Analyse des données
Conceptions non expérimentales a) Contrôles élaborés ou harmonisation	Enquête de grande envergure, recensement, budget nationale ou type d'enquête de la méthode des moindres carrés ordinaires (MCO) sur échantillonnage des bénéficiaires	Enquête de grande envergure et des enquêtes de base sur les ménages, toutes les deux aux points d'indice au moment de la vérification des évènements contemporains	
b) Comparaisons réflexives et différence double	Etude de base et suivi des bénéficiaires	Série chronologique ou enquête par panel sur des bénéficiaires comparables et non comparables	

Tableau 2.1 (Suite)

Méthodologie	Spécification des données		Utilisation d'approche qualitative
	Minimale	Idéale	
c) Contrôle statistique ou variable instrumentale	Données transversales des représentants de la population bénéficiaire avec des variables instrumentales correspondantes	Données transversales et série chronologique des représentants de la population bénéficiaire et de la population non bénéficiaire avec des variables instrumentales correspondantes	

Sources : Adapté de Ezemenari, Rudqvist et Subbarao (1999) et Bamberger

Pour les évaluations qui produiront leurs propres données, il y a les étapes critiques de la conception des instruments de collecte des données, l'échantillonnage, le travail de terrain, la gestion des données et l'accès aux données. Cette section ne décrit pas le processus par étapes de la façon d'entreprendre une enquête, mais fournit plutôt une discussion brève de ces étapes. Une partie de la discussion de cette section, notamment l'échantillonnage et la gestion des données est plus appropriée aux évaluations basées sur la collecte et l'analyse des enquêtes par sondage, d'échelle plus grande, employant des données quantitatives que des évaluations employant des données qualitatives et des petites tailles d'échantillon.

Que faut-il mesurer? Les principaux indicateurs de résultat et d'impact devraient être fixés dans la planification de l'évaluation, probablement comme partie d'une approche par le cadre logique. Pour s'assurer que l'évaluation est capable de mesurer les résultats pendant une période de temps appropriée aux besoins des décideurs, une hiérarchie d'indicateurs pourrait être établie allant des indicateurs d'impact à court terme, comme la fréquentation scolaire, aux indicateurs à long terme comme la réussite de l'élève. Cela assure que même si des impacts finaux ne sont pas pris en compte initialement, les résultats du programme peuvent être évalués. De plus, l'évaluateur doit projeter de mesurer les résultats de l'intervention ainsi que de tenir compte des facteurs exogènes qui peuvent avoir un effet sur ces résultats.

Les directeurs d'évaluation peuvent aussi projeter de mener une évaluation sur plusieurs périodes permettant à des impacts plus immédiats d'être relevés plus tôt, tout en poursuivant toujours des mesures de résultat final. Cela a été fait dans l'évaluation de la Réforme de l'Éducation au Nicaragua dans laquelle l'impact à plus court terme de la réforme sur la participation parentale et l'assiduité de l'étudiant et de l'enseignant

a été établi et les impacts à plus long terme sur la réussite de l'étudiant sont toujours en cours d'évaluation.

On pourrait aussi considérer l'information sur les caractéristiques de la population bénéficiaire non strictement liée à l'évaluation d'impact mais intéressante dans l'analyse comme le niveau de pauvreté ou leur avis sur le programme. De plus, l'évaluateur peut aussi vouloir inclure des mesures de coût afin de procéder à une analyse de la rentabilité ou d'autres évaluations complémentaires non strictement liées à l'évaluation de l'impact.

Le type de conception de l'évaluation choisi pour l'évaluation de l'impact comportera sur des besoins de données. Celles-ci seront spécifiques à la méthodologie, la population concernée, les mesures d'impact et d'autres éléments de l'évaluation. Par exemple, si une approche à variable instrumentale (un des types de stratégies d'harmonisation - comparaison) doit être employée, la variable qui servira comme instrument pour séparer la participation au programme des mesures de résultat doit être identifiée et incluse dans la collecte des données. Ceci a été fait pour l'évaluation de l'impact du Fonds d'investissement social bolivien, dans laquelle la connaissance du fonds social et la présence d'organisations non gouvernementales furent utilisées comme des variables instrumentales dans l'évaluation de l'impact des interventions du fonds social.

Il peut être utile de développer une matrice pour l'évaluation, inscrivant la question d'intérêt à l'étude, les indicateurs de résultat qui seront employés pour évaluer les résultats, la variable et sa source de données. Cette matrice peut alors être employée pour passer en revue des questionnaires et projeter le travail analytique comme cela a été fait dans l'évaluation du Fonds d'investissement social d'urgence au Nicaragua (voir Annexe 6).

Développer les instruments de collecte de données et les approches. Le développement des instruments de collecte des données appropriées qui fourniront les données nécessaires pour répondre aux questions d'évaluation peut être délicat. Cela nécessitera d'impliquer les analystes dans le développement des questions, dans l'enquête pilote et dans l'examen des données. L'implication tant du directeur de terrain que du responsable des données pendant le développement des instruments, aussi bien que du personnel local - de préférence des analystes qui peuvent apporter leur connaissance sur le pays et le programme - peut être décisive pour la qualité de l'information collectée (Grosh et Muñoz 1996). Il est aussi important de s'assurer que les données collectées puissent être désagrégées par genre pour explorer l'impact différentiel des programmes et politiques spécifiques.

Des évaluations quantitatives rassemblent et enregistrent l'information d'habitude sous forme numérique ou en catégories préalablement codées. Avec des évaluations qualitatives l'information est généralement présentée comme un texte descriptif avec une catégorisation ou sans catégorisation. L'information peut inclure les réponses d'un individu à des questions d'interview ouverte, des notes prises pendant des réunions de groupes cibles

ou les observations d'événements par l'évaluateur. Certaines études qualitatives emploient la classification pré codée aussi (Bamberger, 2000). La gamme d'instruments de collecte, des données et leurs forces et faiblesses sont récapitulées dans le tableau 2.2—la technique la plus généralement employée étant les questionnaires.

Les réponses aux questionnaires des enquêtes peuvent être très délicates à concevoir, ainsi il est important de s'assurer que la structure et le format soient appropriés, de préférence établis par un personnel expérimenté. Par exemple, l'utilité des données quantitatives a été souvent sérieusement handicapée pour des raisons mécaniques simples, comme l'incapacité de lier des données d'une source à une autre. Ce fut le cas dans une évaluation d'éducation nationale dans un pays où les données de contexte sur les étudiants, ne pouvaient pas être liées aux résultats de test, ce qui a rendu impossible d'évaluer l'influence des caractéristiques estudiantines sur la performance ou de classer les résultats des tests par âge des étudiants, le genre, le statut socio-économique ou l'histoire éducative.

Pour la collecte des données tant qualitatives que quantitatives, le personnel même expérimenté doit être formé pour collecter les données spécifiques à l'évaluation et toute la collecte des données doit être guidée par un jeu de manuels qui peuvent être employés comme orientation pendant la formation et comme référence pendant le travail de terrain. Selon la complexité de la tâche de la collecte de données, les exemples de cas montrent que la formation peut s'étendre de trois jours à plusieurs semaines.

L'enquête pilote est une étape essentielle parce qu'elle révélera si l'instrument peut sûrement produire les données nécessaires et comment les procédures de la collecte de données peuvent être mises en opération. Le test pilote doit imiter le travail de terrain réel aussi étroitement que possible. Pour cette raison, il est utile d'avoir des programmes de saisie de données prêts au moment du test pilote pour éprouver leur fonctionnalité aussi bien que l'enquête pilote à travers des populations et des secteurs géographiques différents.

Tableau 2.2 Principaux instruments de collecte des données pour l'évaluation d'impact

<i>Technique</i>	<i>Définition et utilisation</i>	<i>Forces</i>	<i>Faiblesses</i>
Etudes de cas	Collecte des renseignements qui résultent en une histoire descriptive ou explicative pouvant aider à répondre aux questions de comment et pourquoi	<ul style="list-style-type: none"> - Peut traiter une variété de preuve à partir des documents, interviews, observation - Peut associer une force explicative quand l'accent est mis sur des institutions, processus, programmes, décisions et événements 	<ul style="list-style-type: none"> - Bonnes études de cas difficiles à réaliser - Nécessite des recherches spéciales et des aptitudes de rédacteur pour être rigoureux - Conclusions des recherches ne sont pas généralisables à la population - Prend du temps - Difficile à transposer
Groupe cible	Discussions ciblées avec des membres de la population concernée qui sont familiers avec des questions pertinentes avant la rédaction d'une série de questions structurées. L'objectif est de comparer des perspectives des bénéficiaires avec des concepts abstraits dans les objectifs de l'évaluation	<ul style="list-style-type: none"> - Avantages similaires aux interviews (ci-dessous) - Particulièrement utile là où l'interaction est désirée - Une manière utile d'identification des influences hiérarchiques 	<ul style="list-style-type: none"> - Peut être coûteux et prendre du temps - Doit être sensible à la combinaison des niveaux hiérarchiques - N'est pas généralisable
Interviews	L'interviewer pose des questions à une ou plusieurs personnes et enregistre les réponses des personnes interrogées. Les interviews peuvent être formels ou informels, face à face ou par téléphone, ou fermés ou ouverts	<ul style="list-style-type: none"> - Les gens et les institutions peuvent expliquer leurs expériences dans leur propre langue et cadre - Flexible et permet à l'interviewer de poursuivre des lignes d'enquêtes non anticipées et examiner les questions en profondeur 	<ul style="list-style-type: none"> - Prend du temps - Peut être coûteux <ul style="list-style-type: none"> - Si l'enquête n'est pas convenablement menée, l'interviewer peut influencer la réponse de la personne interrogée
		<ul style="list-style-type: none"> - Particulièrement utile quand les difficultés de langage sont anticipées 	

Observation	<p>Observe et enregistre la situation dans un registre ou un agenda. Ceci englobe les personnes concernées; ce qui s'est passé; quand, où et comment les événements se sont passés. L'observation peut être directe (l'observateur observe et enregistre), ou participative (l'observateur devient partie intégrante du cadre pour une période de temps). Développe une série de questions d'enquête dont les réponses peuvent être régulièrement codées</p>	<ul style="list-style-type: none"> - Possibilité accrue d'obtenir des données de responsables - Fournit une information descriptive sur le contexte et des changements observés 	<ul style="list-style-type: none"> - La qualité et l'utilité des données dépendent largement des compétences de rédaction et d'observation de l'observateur. - Les conclusions des recherches peuvent être ouvertes à une interprétation - Ne s'applique pas facilement au cadre d'une courte période de traitement du changement
Questionnaires		<ul style="list-style-type: none"> - Peut atteindre un échantillon plus large simultanément - Permet aux personnes interrogées de réfléchir quelques instants avant de répondre - Permet de répondre anonymement - Impose l'uniformité en posant les mêmes questions à toutes les personnes interrogées - Facilite la comparaison et la compilation des données de base 	<ul style="list-style-type: none"> - La qualité des réponses dépend largement de la clarté des questions - Parfois il est difficile de persuader les gens de remplir et retourner le questionnaire - Peut impliquer des activités institutionnelles et des expériences des catégories de gens d'influence prédéterminés
Analyse d'un document écrit	<p>Révision des documents tels que les fichiers des données de base, des données administratives, des matériels de formation et de la correspondance</p>	<ul style="list-style-type: none"> - Peut identifier des questions à examiner plus tard et fournit une preuve pour l'action, le changement et l'impact pour soutenir des perspectives des personnes interrogées - Peut être à moindre coût 	<ul style="list-style-type: none"> - Peut prendre du temps

Source : Adapté de Taschereau (1998)

Echantillonnage. L'échantillonnage est un art le mieux pratiqué par un spécialiste d'échantillonnage expérimenté. La conception n'a pas besoin d'être compliquée mais elle doit être informée par l'expertise du spécialiste de prélèvement d'échantillons dans la détermination des cadres appropriés au prélèvement des échantillons, des tailles et des stratégies de choix (la discussion du tirage des échantillons incluse ici se réfère principalement aux questions liées aux évaluations qui collectent des données quantitatives à partir des échantillons plus grands, statistiquement représentatifs). Le spécialiste d'échantillonnage doit être intégré dans le processus d'évaluation dès les toutes premières étapes pour passer en revue l'information disponible nécessaire pour choisir l'échantillon et déterminer si quelque travail d'énumération sera nécessaire, ce qui pourrait prendre du temps.

Comme avec d'autres parties du travail d'évaluation, la coordination entre le spécialiste de l'échantillonnage et l'équipe d'évaluation est importante. Ceci devient particulièrement critique dans la conduite de comparaisons d'harmonisation parce que la conception de tirage des échantillons devient la base pour l'harmonisation qui est au "cœur" de

34 DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT

la conception de l'évaluation et de la construction des comparaisons contre factuelles. Dans ces cas, le spécialiste du tirage des échantillons doit travailler étroitement avec l'équipe d'évaluation pour développer les critères qui seront appliqués pour harmoniser les groupes de traitement et de comparaison. Par exemple, dans l'évaluation de la réforme de l'autonomie de l'école au Nicaragua, des écoles autonomes étaient stratifiées par le type d'école, l'inscription, la durée de la réforme et l'emplacement des écoles non autonomes, en employant les mêmes stratifications sauf la durée de la réforme. Cela peut être facilité par la présence d'un membre d'équipe responsable du travail de la collecte de données qui assiste le spécialiste du prélèvement des échantillons dans l'obtention de l'information nécessaire y compris des données sur les indicateurs de résultats choisis pour les calculs de masse (une estimation de la taille type nécessaire pour évaluer la signification statistique entre deux groupes), une liste de la population concernée pour le choix type et des détails sur les caractéristiques des groupes de traitement et de comparaison potentiels, importantes pour le processus de choix de l'échantillon.

Il y a beaucoup d'écarts entre les coûts et la précision de l'échantillonnage qui devraient être éclaircis pendant que la structure de l'échantillonnage se développe. Par exemple en conduisant un échantillon sur deux ou trois niveaux, on réduira les dépenses tant du tirage des échantillons que du travail de terrain, mais les erreurs de tirage des échantillons, et donc la précision des évaluations, seront augmentées.

Une fois que les variables de résultats et la population concernée ont été déterminés par l'équipe d'évaluation, une première démarche pour le spécialiste du tirage des échantillons serait de déterminer le calcul de masse (voir Valadez et Bamberger 1994, pp. 382-84, pour une discussion du processus). Puisque le calcul de masse peut être exécuté en employant seulement une mesure de résultat et que les évaluations en considèrent souvent plusieurs, quelques décisions stratégiques devront être prises au sujet de quel indicateur de résultat employer en concevant l'échantillon.

Après le développement de la stratégie et de la structure de l'échantillonnage, le spécialiste du tirage des échantillons doit aussi être associé à la sélection de l'échantillon pour le travail de terrain et le test pilote pour assurer que ce dernier n'est pas conduit dans un secteur qui sera intégré dans l'échantillon pour le travail de terrain. Le travail de terrain initial sera souvent nécessaire comme une partie de la procédure de choix de l'échantillon. Par exemple, un processus d'énumération sera nécessaire s'il n'y a pas de cartes à jour des unités nécessaires pour l'échantillon (les ménages, les écoles et ainsi de suite) ou si une certaine population cible, comme des enfants sous alimentés, doit être identifiée afin de pouvoir être choisie pour le but de l'évaluation.

Une fois que le travail de terrain est achevé, le spécialiste du tirage des échantillons doit fournir l'assistance en déterminant les poids du tirage des échantillons pour calculer les facteurs d'expansion et corriger les erreurs d'échantillonnage et l'absence de réponse (Grosh et Muñoz 1996 fournit une discussion détaillée des procédures d'échantillonnage comme une partie du travail d'enquête sur le ménage. Kish 1965 est considéré comme l'un des manuels standards

DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT 35

dans le domaine de l'échantillonnage). Et finalement, le spécialiste du tirage des échantillons devrait produire un document détaillant la stratégie de l'échantillonnage y compris (a) l'étape de la conception de l'échantillonnage, les calculs de masse employant les variables d'impact, la détermination des erreurs d'échantillonnage et des tailles, l'utilisation de la stratification pour analyser les populations cibles, (b) à partir de l'étape de sélection de l'échantillon, un profil des étapes du prélèvement des échantillons et des procédures de choix, (c) à partir de l'étape du travail de terrain à préparer pour l'analyse, le rapport entre la taille de l'échantillon et la population dont il a été choisi les taux d'absence de réponse et d'autres renseignements employés pour informer les poids, et l'utilisation des données d'évaluation. Ce document peut être employé pour conserver les rapports du projet d'évaluation et devrait être intégré aux données chaque fois qu'elles sont distribuées pour aider à guider les analystes dans l'utilisation des données d'évaluation.

Questionnaires. L'élaboration du questionnaire est importante pour la validité de l'information collectée. Il y a quatre types généraux d'information nécessaire pour une évaluation d'impact (Valadez et Bamberger 1994). Ce sont :

- La classification des données nominales avec des personnes interrogées classées selon qu'elles sont participantes au projet ou appartiennent au groupe témoin
- L'exposition aux variables de traitement enregistrant non seulement les services et bénéfices reçus mais aussi la fréquence, la quantité et la qualité - évaluer la qualité peut être très difficile;
- Les variables de résultat pour mesurer les effets d'un projet y compris des produits immédiats, les résultats durables ou la prestation continue de services sur une longue période et des impacts de projet comme des revenus améliorés et l'emploi ;

- Les variables incidentes qui affectent la participation à un projet ou le type d'impact produit, comme l'individu, le ménage, ou des caractéristiques de la communauté - ces variables peuvent être importantes pour l'exploration des distorsions.

La manière dont on pose les questions, aussi bien que l'ordre des questions, sont tout aussi importants dans la production d'information fiable. Un exemple approprié est la mesure du bien-être qui serait nécessaire pour mesurer l'impact direct d'un projet sur la réduction de la pauvreté. Demander aux individus le niveau de leurs revenus ne fournirait pas nécessairement des résultats précis sur leur niveau de bien-être économique. Comme discuté dans la littérature sur la mesure du bien-être, des questions sur les dépenses, la composition des ménages, les actifs, les cadeaux et les envois de fonds et la valeur imputée de l'alimentation produite dans le pays et les logements occupés par les propriétaires sont généralement

36 DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT

employés pour saisir la vraie valeur du bien-être des ménages et des individus. Le rappel du temps, utilisé parmi les éléments de dépense ou l'ordre dans lequel on pose ces questions, peut significativement affecter la validité de l'information collectée.

Parmi les éléments notés pour un bon questionnaire, il y a : le garder court et axé sur des questions importantes, s'assurer que les instructions et les questions soient claires, limiter les questions à celles nécessaires pour l'évaluation et inclure une option "pas d'opinion" pour les questions fermées afin de s'assurer des données fiables et utiliser des procédures judicieuses pour administrer le questionnaire qui peut en effet être différent pour des enquêtes quantitatives et qualitatives.

Problèmes du travail de terrain. Le travail avec le personnel local ayant une grande expérience dans la collecte de données similaires à celles nécessaires pour l'évaluation peut énormément faciliter les opérations du travail de terrain. Non seulement ce personnel peut fournir l'information nécessaire sur le territoire géographique à couvrir mais leur connaissance peut aussi être critique pour développer des normes employées dans la localisation et l'approche des informateurs. Le personnel de terrain dont l'expertise se situe dans un secteur autre que celui demandé pour l'effort d'évaluation peut présenter des problèmes comme ce fut le cas dans une évaluation de l'éducation au Nicaragua qui a employé une société spécialisée dans le sondage d'opinion publique pour mener une enquête sur les écoles et les ménages. L'expertise qui avait permis à cette société d'acquérir une excellente réputation basée sur sa prévision précise des résultats d'élection incertains n'était pas utile pour savoir comment approcher des écoliers ou fusionner des jeux de données quantitatives. Le manque d'expertise a créé des problèmes substantiels de mise en oeuvre de l'enquête qui ont nécessité des semaines d'action rectificative par une équipe commune du ministère de l'éducation nationale et la Banque Mondiale.

Le type de personnel nécessaire pour la collecte des données sur le terrain variera selon les objectifs et le centre d'intérêt de l'évaluation. Par exemple, une évaluation d'impact quantitative d'un programme de nutrition pourrait nécessiter l'intégration d'un spécialiste de l'anthropométrie, pour collecter des mesures de hauteur pour poids, comme membre de l'équipe d'enquête tandis que l'évaluation d'impact d'une réforme éducative intégrerait très probablement le personnel spécialisé dans l'application des tests de prestations académiques. La plupart des enquêtes quantitatives nécessiteront au moins un responsable d'enquête, un responsable des données, un responsable du travail de terrain, des superviseurs de terrain, des interviewers, des opérateurs de saisie de données et des chauffeurs. Selon l'approche qualitative utilisée, le personnel de terrain peut être similaire, à l'exception des opérateurs de saisie de données. Les compétences des interviewers cependant, seraient tout à fait différentes avec des interviewers qualifiés nécessitant une formation spécialisée, particulièrement pour des groupes cibles, l'observation directe et ainsi de suite.

Il faut se rappeler trois autres préoccupations en projetant des opérations d'enquête. D'abord, il est important de prendre en considération des événements temporels

DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT 37

qui peuvent affecter le succès opérationnel du travail de terrain et la validité externe des données collectées, comme le calendrier de l'année scolaire, les vacances, les saisons pluvieuses, les temps de la moisson ou des migrations de masse. Deuxièmement, il est crucial pour les instruments de collecte de données de tests pilotes, même s'ils sont des adaptations d'instruments qui ont été employés précédemment, à la fois de tester la qualité de l'instrument en vue de produire les données demandées et de familiariser le personnel de terrain avec la dynamique du processus de la collecte de données. Les enquêtes pilotes peuvent aussi servir comme une base de preuves par le choix de l'équipe centrale du personnel de terrain pour effectuer l'enquête proprement dite. Beaucoup de collecteurs de données expérimentés commenceront avec 10 à 20 pour cent de plus de personnel dans l'enquête pilote que celui qui sera utilisé dans le travail de terrain réel et choisira ensuite les meilleurs pour former les équipes réelles de la collecte de données. Finalement, les communications sont essentielles pour des opérations sur le terrain. Par exemple, si les conditions locales permettent leur utilisation, le travail de terrain peut être renforcé en équipant les superviseurs avec des téléphones mobiles pour qu'ils puissent rester en contact avec le responsable d'enquête, le responsable des enquêtes sur le terrain et d'autres personnels pour répondre aux questions et les tenir informés de l'état d'avancement.

Gestion des données et accès. Les objectifs d'un bon système de gestion des données devraient être d'assurer l'opportunité et la qualité des données d'évaluation. L'opportunité dépendra de l'intégration autant que possible entre la collecte de données et leur traitement pour que les erreurs puissent être vérifiées et corrigées avant la fin du travail de terrain. La qualité des données peut être assurée en faisant des contrôles de cohérence pour évaluer la validité interne des données collectées aussi bien pendant qu'après que les données ont été saisies et en assurant que la documentation

appropriée soit disponible pour les analystes qui utiliseront les données. La documentation doit résister à deux types d'information : (a) l'information nécessaire pour interpréter les données incluant des guides, des dictionnaires de données, des guides pour les variables et toutes traductions nécessaires et (b) l'information nécessaire pour conduire l'analyse qui est souvent contenue dans un document d'information de base qui contient une description du centre d'intérêt et de l'objectif de l'évaluation, des détails sur la méthodologie de l'évaluation, des résumés ou copies des instruments de collecte des données, l'information sur l'échantillon, une discussion du travail de terrain et des directives pour l'utilisation des données.

On recommande que les données produites par des évaluations soient rendues disponibles étant donné le caractère de bonne valeur pour les besoins d'évaluation. Un travail de suivi complémentaire pour évaluer des impacts à long terme par une équipe autre que celle qui a effectué le travail d'évaluation original. Pour faciliter le processus du partage des données, au début de l'évaluation, une politique d'accès ouvert aux données doit être convenue et signée, établissant des normes et des responsabilités de la diffusion des données. Une politique d'accès ouvert aux données met une charge supplémentaire sur la bonne documentation des données et la protection

38 DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT

de la confidentialité des enquêtés. Si les données de l'équipe sont collectées des mêmes enquêtés sur une période de temps par des agences différentes, les enquêtés devront être identifiés pour mener le travail complémentaire. Cette exigence doit être équilibrée par rapport aux normes de confidentialité qui accompagnent généralement toute recherche du secteur social. Une solution possible est de rendre les données anonymes du fichier disponibles pour tous les analystes concernés mais de demander aux chercheurs intéressés par la conduite du travail complémentaire d'entrer en contact avec l'agence responsable des données pour obtenir l'inscription des données dans l'échantillon donnant ainsi une occasion à l'agence d'assurer le contrôle de la qualité du travail à venir grâce au contact avec les chercheurs voulant l'effectuer.

Analyse, Rapport et Diffusion

Comme avec d'autres étapes du processus d'évaluation, l'analyse des données d'évaluation, quantitatives ou qualitatives, exige la collaboration entre les analystes, les producteurs de données et les décideurs pour clarifier des questions et s'assurer des résultats de qualité en temps opportun. Des problèmes avec le peaufinage et l'interprétation des données surgiront presque sûrement pendant l'analyse et nécessiteront la contribution des divers membres de l'équipe.

Certaines des techniques et des défis à effectuer l'analyse quantitative basée sur des méthodes statistiques sont incluses dans le chapitre 3. Il existe aussi beaucoup de techniques pour analyser des données qualitatives (voir Miles et Huberman 1994). Bien qu'une discussion détaillée de ces méthodologies soit hors de la portée de ce manuel, deux méthodologies généralement employées pour l'évaluation d'impact sont mentionnées - l'analyse de contenu et l'analyse de cas (Taschereau 1998).

L'analyse de contenu est employée pour analyser des données tirées d'interviews, d'observations et de documents. Dans l'examen des données, l'évaluateur développe un système de classification pour les données, l'organisation de l'information basée sur (a) les questions d'évaluation par lesquelles l'information a été collectée, (b) comment le matériel sera employé, et (c) le besoin de renvoyer l'information. Le codage des données peut être tout à fait complexe et peut nécessiter beaucoup d'hypothèses. Une fois qu'un système de classification a été mis en place, la phase de l'analyse commence, qui est aussi un processus difficile. Cela implique la recherche de modèles dans les données et un déplacement au-delà de la description vers le développement d'un ensemble de processus de programme, des résultats et des impacts. Cela se fait mieux avec la participation des membres de l'équipe. De nouveaux programmes informatiques, ethnographiques et linguistiques sont aussi maintenant disponibles, conçus pour soutenir l'analyse des données qualitatives.

L'analyse de cas est basée sur des études de cas conçues pour une étude approfondie d'un groupe particulier ou d'un individu. Le niveau élevé du détail peut fournir une information riche pour évaluer l'impact du projet. Les processus de la collecte et

DEMARCHES CLES DANS LA CONCEPTION ET LA MISE EN ŒUVRE DES EVALUATIONS D'IMPACT 39

de l'analyse des données sont réalisés simultanément, puisque les évaluateurs font les observations pendant qu'ils collectent les informations. Ils peuvent alors développer et évaluer des explications et relier les renseignements critiques.

Qu'on analyse l'information quantitative ou qualitative, d'autres leçons générales liées à l'analyse, le rapport et la diffusion peuvent aussi être tirées des exemples de cas de l'Annexe 1.

Premièrement, l'analyse prend généralement plus de temps que prévu, particulièrement si les données ne sont pas aussi propres ou accessibles au début de l'analyse, si les analystes n'ont pas l'expérience du type de travail d'évaluation, ou s'il y a un accent mis sur le renforcement des capacités par le travail fait en collaboration. Dans l'examen des études de cas traitées pour cet article, l'analyse la plus rapide a pris environ un an après la production des données et l'analyse la plus longue a pris près de deux ans. Le cas du chapitre 3 illustre certaines des nombreuses étapes intégrées dans l'analyse et la raison pour laquelle elle peut prendre plus de temps que prévu.

Deuxièmement, l'évaluateur doit projeter de sortir plusieurs produits comme résultats du travail analytique, gardant à l'esprit deux éléments. Le premier est de s'assurer du choix du temps des productions qui doit se situer autour d'événements clés : temps de prise de décisions sur l'avenir du projet comme les revues à mi-parcours ; les élections, la fermeture d'une phase pilote. Le second est l'audience à laquelle les résultats sont destinés. Les produits doivent être différenciés selon l'audience pour laquelle ils sont conçus, y compris les décideurs du gouvernement, les responsables de programmes, les donateurs, le grand public, les journalistes et les universitaires.

Troisièmement, les produits auront la meilleure pertinence politique s'ils contiennent des recommandations claires et pratiques découlant de l'analyse d'impact. Celles-ci peuvent être séparées en priorités à court et à long terme et si possible, inclure des incidences budgétaires. Les décideurs seront enclins à chercher la fin des recommandations.

Finalement, les rapports doivent être prévus comme partie intégrante d'une stratégie de diffusion élargie qui peut inclure des présentations pour des audiences diverses, des communiqués de presse,

des informations aux informateurs et la diffusion sur l'Internet. Une telle stratégie de diffusion doit être incluse au stade initial du processus de planification pour s'assurer qu'il en est tenu compte dans le budget et que les résultats atteindront l'audience prévue.

Chapitre 3

Application de Méthodes Analytiques pour l'Evaluation d'Impact : Une Etude de Cas *

Cette étude de cas est basée sur un programme de lutte contre la pauvreté hypothétique PROSCOL, qui fournit des transferts d'argent à des familles pauvres avec des enfants d'âge scolaire dans une région d'un pays en développement. Le cas est destiné à illustrer les étapes analytiques impliquées dans l'exécution d'une évaluation d'impact et les options auxquelles un analyste peut faire face avec le processus applicable à n'importe quel type de programme de lutte contre la pauvreté. En explorant comment procéder pour évaluer l'impact du programme, l'analyste politique commet plusieurs erreurs le long du parcours, cherchant des informations sur des sujets spécifiques à partir des compétences spécialisées des collègues-un statisticien, un économiste, un professeur d'économétrie et un sociologue.

Parmi les étapes analytiques que l'analyste parcourt dans ce cas, il y a l'identification des questions à traiter dans l'évaluation d'impact, l'évaluation des ressources de données, un premier aperçu sur les données, une compréhension des distorsions des variables de contrôle, une information sur les revenus prévus, la combinaison de variables, la compréhension de l'importance de l'exogénéité, l'exploration des meilleures façons de former un groupe de comparaison (harmonisation du score de propension), information sur les distorsions et examens de ce qui pourrait avoir été fait avec une enquête de données de base (des différences doubles), l'utilisation des variables instrumentales, l'essai des diverses méthodologies, l'intégration des ressources venant du terrain et la planification du travail à venir.

Description du programme hypothétique PROSCOL

Le programme PROSCOL identifie des familles éligibles à la participation en utilisant les diverses manifestations de pauvreté qui comprennent le nombre des personnes dans les ménages, l'éducation du chef de famille et les diverses caractéristiques du logement. PROSCOL paye une somme fixée par enfant d'âge scolaire à tous les ménages choisis à condition que les enfants aient un taux d'assiduité de 80 pour cent ce qui doit être vérifié par une note de l'école. Les ménages doivent garder leurs enfants à l'école jusqu'à 18 ans.

Ce programme initié il y a 12 mois est financé par la Banque Mondiale et fonctionne en dehors du Ministère du Développement Social. Dans un effort

* Ce chapitre a été développé dans les détails dans un document préparé par Martin Ravallion, *The Mystery of the Vanishing Benefits : Ms. Speedy Analyst's Introduction to Evaluation*, Policy Research Working Paper No. 2153, 1999.

pour évaluer l'impact du PROSCOL sur la pauvreté en vue d'aider à déterminer si le programme doit être généralisé pour inclure le reste du pays ou être abandonné, la Banque Mondiale a demandé une évaluation d'impact au Ministère des Finances. La demande était adressée au Ministère des Finances afin d'aider à assurer une évaluation indépendante et aider à développer la compétence pour ce type d'évaluation dans une unité centrale du Gouvernement – près du centre d'allocations budgétaires.

Identifier les Questions à Traiter dans l'Evaluation d'Impact

La première étape pour l'analyste au Ministère des Finances assigné à la tâche d'effectuer l'évaluation PROSCOL consiste à clarifier les objectifs du projet à examiner dans l'évaluation de l'impact. Le projet a deux buts politiques : les transferts d'argent visent à réduire la pauvreté actuelle, et en insistant pour que les destinataires de transferts gardent leurs enfants à l'école ; le programme vise à réduire la pauvreté dans l'avenir en élevant les niveaux d'éducation parmi la population actuelle des enfants pauvres. Deux éléments d'information seraient donc nécessaires pour évaluer l'impact. D'abord, les transferts d'argent vont – ils principalement aux familles à faible revenu ? Et ensuite quelle augmentation des taux d'inscription scolaire le programme entraîne-t-il ?

Evaluer les Ressources de Données

Pour effectuer l'évaluation, l'analyste a deux ressources principales. La première est un rapport basé sur des interviews qualitatives avec des administrateurs de programme et les groupes cibles de participants. Il n'est pas évident, cependant, que ceux qui étaient interviewés soient représentatifs des participants au programme PROSCOL et si leur pauvreté était plus prononcée par rapport à ceux qui n'ont pas été choisis pour le programme et n'ont pas été interviewés. Le rapport indique que les enfants sont allés à l'école mais il est possible qu'ils aient pu aussi aller à l'école si le programme n'avait pas existé. Bien que ce rapport soit un début important, il ne dit pas à l'analyste comment les participants du PROSCOL sont pauvres et quel impact le programme a sur la scolarisation. La deuxième ressource est une enquête nationale indépendante récente, effectuée par le Bureau des Statistiques du pays sur les ménages dénommée Enquête des niveaux de vie (ENV). La ENV a porté sur un échantillon aléatoire de 10.000 ménages et a interrogé sur les revenus des ménages par source, l'emploi, les dépenses, l'état de santé, les résultats d'éducation et les caractéristiques démographiques et autres de la famille. L'enquête avait incorporé une question pour savoir si l'échantillon de ménage avait participé ou non à PROSCOL et une ligne pour l'argent reçu du PROSCOL dans la liste des sources de revenu.

Premier Aperçu sur les Données

L'analyste continue par obtenir les données brutes de ENV pour se concentrer à évaluer des bénéficiaires du programme. L'analyste emploie un logiciel statistique comme SPSS ou SAS pour réaliser une tabulation croisée de la moyenne des sommes d'argent reçus du PROSCOL par les "déciles" de ménages (les déciles sont formés en classant tous les ménages de l'échantillon selon leur revenu par personne). Dans le calcul de ce dernier, l'analyste décide d'extraire toutes les sommes d'argent reçues du PROSCOL comme une bonne mesure de revenu en l'absence du programme avec l'intention d'identifier qui a gagné par rapport à son revenu avant l'intervention.

La tabulation croisée suggère que les transferts d'argent au titre du programme soient tout à fait bien allés aux pauvres. Selon le seuil de pauvreté officiel du pays, environ 30 pour-cent de la population dans le Nord-ouest sont pauvres. A partir du tableau, les calculs montrent que les 30 pour – cent les plus pauvres de l'échantillon d'enquête reçoivent 70 pour – cent des transferts de PROSCOL. A première vue cela semble être un résultat positif.

La question suivante concerne l'impact sur la scolarisation. Cela est examiné à travers une tabulation croisée des taux d'inscription scolaire moyens des tranches d'âge diverses pour des familles PROSCOL en comparaison avec les familles non PROSCOL. Cette comparaison ne révèle presque aucune différence entre les deux ; le taux d'inscription moyen pour des enfants âgés de 6 à 18 ans est environ 80 pour-cent dans les deux cas. L'analyste calcule alors les années moyennes d'études à chaque âge et les résultats sont représentés par graphique séparément pour des familles PROSCOL et des familles non PROSCOL. Cela montre que les deux figures ne sont pas identiques mais elles sont très proches. A cette étape, l'analyste se demande s'il n'y avait vraiment aucun impact sur la scolarisation ou si l'approche est fausse.

Comprendre les Distorsions

Avec cette incertitude, l'analyste cherche ensuite l'assistance d'un statisticien confirmé pour explorer pourquoi les résultats démontrent que les enfants PROSCOL ne sont probablement pas plus scolarisés que les enfants non PROSCOL. Le statisticien formule une hypothèse que les résultats peuvent présenter une sérieuse distorsion. Pour évaluer l'impact du programme, nous devons savoir ce qui serait arrivé sans le programme. Cependant l'analyste ne l'a pas expliqué ; au lieu de cela les familles non PROSCOL sont utilisées comme groupe de comparaison pour déduire ce que la scolarisation des participants PROSCOL aurait été si le programme n'avait pas existé.

Autrement dit, P_i dénote la participation PROSCOL de l'enfant i . Cela peut prendre deux valeurs possibles à savoir $P_i = 1$ si l'enfant participe à PROSCOL et $P_i = 0$ s'il ou elle ne participe pas. Si le $i^{\text{ème}}$ enfant ne participe pas alors son niveau de scolarité est donc S_{0i} , soit S pour la scolarisation et i pour l'enfant i lorsque $P = 0$ si l'enfant ne participe pas. Si l'enfant participe, alors sa scolarisation est S_{1i} . Son

gain en scolarisation dû à PROSCOL est $S_{1i} - S_{0i}$. Le gain du $i^{\text{ème}}$ enfant qui participe ($P = 1$) est alors :

$$G_i = S_{1i} - S_{0i} \mid P_i = 1.$$

La barre signifie "étant donné que" ou "à condition que" et est nécessaire pour dire clairement que le calcul est le gain pour un enfant qui a réellement participé au programme. Si on veut connaître le gain moyen, c'est simplement la moyenne de tous les G_i , qui donne le gain moyen type de scolarisation parmi tous ceux qui ont participé à PRO SCOL. Tant que cette moyenne est calculée correctement (en utilisant les poids types appropriés de l'enquête), il fournira une évaluation exacte du vrai gain moyen. Ce dernier est "la valeur attendue " de G et il peut être écrit :

$$G = E(S_{1i} - S_{0i} \mid P_i = 1).$$

C'est une autre façon de dire "moyen". Cependant, cela n'a pas besoin d'être exactement égal à la moyenne calculée à partir des données types, entendu qu'il y aura des erreurs d'échantillonnage. Dans l'exposé sur l'échantillonnage de l'évaluation,

$E(S_{1i} - S_{0i} \mid P_i = 1)$ est parfois appelé "l'effet de traitement" ou "l'effet de traitement moyen sur le traité". Dans ce cas PROSCOL est considéré comme le traitement.

Le statisticien fait remarquer à l'analyste qu'elle n'a pas calculé G , mais plutôt la différence de scolarisation moyenne entre des enfants des familles PROSCOL et ceux des familles non PROSCOL. C'est l'estimation de l'échantillon en :

$$D = E(S_{1i} \mid P = 1) - E(S_{0i} \mid P = 0).$$

Il y a une identité simple liant D à G , à savoir :

$$D = G + B.$$

Ce terme "B" est la distorsion dans l'évaluation et est donné par :

$$B = E(S_{0i} \mid P_i = 1) - E(S_{0i} \mid P_i = 0).$$

Autrement dit, la distorsion est la différence attendue dans la scolarisation sans PROSCOL entre les enfants qui ont réellement participé au programme et ceux qui n'y ont pas participé. Cette distorsion pourrait être corrigée si $E(S_{0i} \mid P_i = 1)$ étaient connus, mais il n'est pas possible même de pouvoir l'estimer. On ne peut pas observer ce que la scolarisation aurait été pour des enfants qui ont

réellement participé au programme PROSCOL, s'ils n'avaient pas participé ; cela constitue des données manquantes – appelées aussi moyenne "contre factuelle".

Cette distorsion présente une préoccupation principale. En l'absence du programme PROSCOL des parents peuvent envoyer leurs enfants à l'école moins que d'autres parents. S'il en est ainsi alors il y aura une distorsion dans le calcul. Retournant aux questions d'évaluation originale, nous nous sommes intéressés au surplus de scolarisation dû au programme PROSCOL. Vraisemblablement, ceci affecte seulement des familles qui participent en réalité. En d'autres termes, nous avons besoin de savoir combien de personnes devraient être scolarisées s'il n'y avait pas le programme. S'il n'y a aucune distorsion, donc le surplus de scolarisation au titre du programme est la différence en scolarisation moyenne entre ceux qui ont participé et ceux qui n'ont pas participé au programme. Ainsi la distorsion surgit s'il y a une différence en scolarisation moyenne entre des parents PROSCOL et des parents non PROSCOL en l'absence du programme.

Pour éliminer cette distorsion, la meilleure approche serait d'attribuer le programme de façon aléatoire. Alors les participants et les non participants auront la même chance de scolarisation en l'absence du programme c'est-à-dire $E(S_{0i} | P_i = 1) = E(S_{0i} | P_i = 0)$. La scolarisation des familles non participantes indiquera alors correctement la comparaison contre factuelle c'est-à-dire la scolarisation que nous aurions observée pour des participants s'ils n'avaient pas eu accès au programme. En effet la nomination aléatoire égalisera toute la distribution, pas seulement les moyennes. Il y aura une distorsion suite à une erreur d'échantillonnage mais pour des échantillons assez importants on peut sans risque supposer que chaque différence statistiquement significative dans la distribution de la scolarisation entre des participants et des non participants est attribuable au programme.

Dans la conception actuelle du programme, il est clair que la participation n'est pas aléatoire. En effet, ce serait une critique sérieuse du programme PROSCOL de constater qu'elle l'est. Le fait même de son ciblage intentionnel des familles pauvres, qui sont présumées envoyer moins vraisemblablement leurs enfants à l'école créerait la distorsion.

Cela soulève la question que si PROSCOL marche bien alors nous devons nous attendre à ce que des participants aient une scolarisation pire en l'absence du programme. Alors $E(S_{0i} | P_i = 1) < E(S_{0i} | P_i = 0)$ et le calcul original des analystes sous-estimera le gain du programme. Nous pouvons trouver un minimum de profit ou non bien que le programme marche en réalité bien.

L'analyste se rend maintenant compte que l'ampleur de cette distorsion pourrait être énorme. Supposons que des familles pauvres envoient leurs enfants travailler plutôt que d'aller à l'école ; qu'elles sont pauvres et ne peuvent pas emprunter facilement, elles ont besoin d'argent frais supplémentaire maintenant. Des familles non pauvres envoient leurs enfants à l'école. Le programme choisit des familles pauvres qui envoient alors leurs enfants à l'école. Une différence négligeable existe en scolarisation moyenne entre des familles PROSCOL et des familles non PROSCOL ; en effet, on espère que $E(S_{1i} | P_i = 1) = E(S_{0i} | P_i = 0)$. Mais l'impact du programme est positif et est donné par $E(S_{0i} | P_i = 0) - E(S_{0i} | P_i = 1)$. La défaillance du ciblage intentionnel du programme axé sur les pauvres à prendre en compte pourrait bien avoir conduit à une

sous – estimation des bénéfiques du PROSCOL provenant de la comparaison par l'analyste de la scolarisation moyenne entre des familles PROSCOL et des familles non PROSCOL.

Information sur les revenus du passé

L'analyste montre ensuite les résultats de sa tabulation croisée des montants reçus du programme PROSCOL par rapport au revenu à un autre collègue, un économiste, du Ministère des Finances. L'économiste soulève une préoccupation majeure- que les gains octroyés aux pauvres du programme PROSCOL ont été clairement su-estimés parce que le revenu passé a été ignoré. Les enfants doivent aller à l'école si la famille doit percevoir le transfert ; ainsi ils ne pourront pas travailler, soit dans l'entreprise familiale soit sur le marché du travail. Par exemple, des enfants âgés de 15 à 18 ans peuvent gagner les deux-tiers ou plus du salaire de l'adulte en agriculture ou en construction. Des familles PROSCOL perdront ce revenu du travail de leurs enfants. Ce revenu abandonné devrait être pris en compte quand des gains de revenu net provenant du programme sont calculés. Et ce revenu net devrait être déduit, non du transfert brut, pour faire ressortir le revenu avant intervention. Il importerait aussi de déterminer la pauvreté de la famille en l'absence du transfert PROSCOL. Le présent tableau, par conséquent, pourrait considérablement exagérer les gains du programme pour le pauvre.

L'analyste se demande pourquoi elle devrait déduire le revenu du travail de l'enfant, avec l'hypothèse que le travail des mineurs n'est pas recommandable. L'économiste relève qu'elle devrait considérer les gains dus à la réduction du travail de l'enfant dont le gain principal est la scolarisation complémentaire et par conséquent des revenus futurs plus élevés, pour des familles présentement pauvres. L'analyste a produit des tableaux qui reflètent les deux principales voies de réduction de la pauvreté par PROSCOL : en augmentant les revenus présents des pauvres et en augmentant leurs revenus futurs. L'impact sur le travail des enfants est important dans les deux sens mais dans des directions opposées; ainsi PROSCOL fait face à un dilemme.

Cela démontre pourquoi il est important d'avoir une bonne estimation de l'impact sur la scolarisation; seulement alors il sera possible de déterminer le manque à gagner. Il est par exemple possible que le temps supplémentaire à l'école vienne en déduction du temps de travail.

En considérant la deuxième tabulation croisée, la principale préoccupation soulevée par l'économiste est qu'il n'y a pas de prime pour tous les autres déterminants de la scolarisation en dehors de la participation au PROSCOL. L'économiste suggère de mener une régression des années de scolarisation sur un jeu de variables de contrôle aussi bien que de savoir si la famille de l'enfant est couverte par PROSCOL ou non. Pour l'enfant i^{me} dans l'échantillon nous avons :

$$S_i = a + bP_i + cX_i + ?_i$$

Ici a , b et c sont des paramètres ; x représente les variables de contrôle comme l'âge de l'enfant, l'éducation de la mère et du père, la taille et la composition démographique des ménages et les caractéristiques scolaires ; et ϵ est un résiduel qui comprend d'autres déterminants de scolarisation et d'erreurs de mesure. La valeur estimée de b vous donne l'impact de PROSCOL sur la scolarisation.

Notez que si la famille de l'enfant i participe à PROSCOL donc $P = 1$ et ainsi sa scolarisation sera $a + b + cX_i + \epsilon_i$. S'il ne participe pas, donc $P = 0$ et ainsi sa scolarisation sera $a + cX_i + \epsilon_i$. La différence entre les deux est le gain en scolarisation en raison du programme, qui est juste b .

Addition de variables de contrôle

Comme suggéré, l'analyste opère une régression avec ou sans les variables de contrôle. Quand elle est rentrée sans les variables les résultats montrent que la valeur estimée, de b n'est pas significativement différente de 0 (en utilisant le test standard-t donné par le paquet statistique). Ces résultats présentent une configuration semblable aux premiers résultats, prenant la différence des moyennes entre les participants et les non participants – suggérant que PROSCOL n'a pas d'impact sur la scolarisation. Cependant, quand plusieurs variables de contrôle sont intégrés à la régression, il y a un coefficient positif et significatif sur la participation au programme PROSCOL. Le calcul montre qu'à 18 ans le programme a ajouté deux ans à la scolarisation.

L'analyste se demande pourquoi ces variables de contrôle font une telle différence. Et les bons contrôles sont-ils faits ? Elle visite ensuite son ancien professeur d'économétrie et lui montre ses régressions. Sa première préoccupation liée à la régression de la scolarisation sur P et X est que cela ne permet pas à l'impact du programme de varier avec x ; l'impact est le même pour chacun, ce qui ne semble pas très probable. Les parents plus scolarisés devraient être plus nombreux vraisemblablement à envoyer les enfants à l'école, donc les bénéficiaires dont ils jouiront du programme PROSCOL seront inférieurs. Pour permettre aux bénéficiaires de varier avec X , laissons les études moyennes de non participants, soit $a_0 + c_0 X_i$ tandis que celle des participants est $a_1 + c_1 X_i$, ainsi le niveau observé de scolarisation est

$$S_i = (\alpha_1 + c_1 X_i + \epsilon_{1i}) P_i + (a_0 + c_0 X_i + \epsilon_{0i}) (1 - P_i)$$

où ϵ_{0i} et ϵ_{1i} sont des erreurs aléatoires, chacun avec des moyennes de zéro et sans corrélation avec X . Pour évaluer ce modèle il est nécessaire d'ajouter un terme complémentaire pour les effets d'interaction entre la participation au programme et des caractéristiques observées à la régression déjà menée.

Ainsi la régression augmentée est :

$$S_i = a_0 + (a_1 - a_0) P_i + C_0 X_i + (C_1 - C_0) P_i X_i + \epsilon_i$$

où $\gamma_i = \gamma_1 P_i + \gamma_0 (1 - P_i)$. Alors $(a_1 - a_0) + (C_1 - C_0) X$ est l'impact moyen du programme pour n'importe quelle valeur donnée de X . Si la moyenne X dans l'échantillon des participants est utilisée, elle donnera donc le gain moyen provenant du programme.

Compréhension de l'importance de l'exogénéité

Une deuxième préoccupation relevée par le professeur d'économétrie est de savoir comment la régression a été estimée. Dans l'utilisation du contrôle de régression dans le "paquet statistique", les moindres carrés ordinaires (MCO), il y a une source de préoccupation parce que les estimations MCO des paramètres seront biaisées même avec de grands échantillons à moins que les variables de droite ne soient exogènes. L'exogénéité signifie que les variables de droite sont déterminées indépendamment des choix de scolarisation et ne sont ainsi pas en corrélation avec le terme d'erreur dans la régression de la scolarisation. Parce que la participation au programme à été intentionnellement ciblée, la participation au programme PROSCOL n'est pas exogène. Cela peut affecter le calcul de l'impact du programme comme suit : l'équation pour les années de scolarisation est :

$$S_i = a + bP_i + cX_i + \gamma_i.$$

La valeur $a + b + cX_i + \gamma_i$ a été utilisée comme estimation de la scolarisation du ménage i quand il participe au programme PROSCOL alors que $a + cX_i + \gamma_i$ a été utilisée pour estimer la scolarisation s'il ne participe pas. Ainsi la différence, b , est le gain provenant du programme. Cependant, en faisant ce calcul, l'hypothèse est que γ_i était le même de toute façon. Autrement dit, l'hypothèse était que γ_i était indépendant de P , ce qui affecterait le calcul de l'impact du programme.

Cela met en évidence la distorsion relative au placement du programme non aléatoire qui peut aussi affecter l'estimation basée sur le modèle de régression suggéré auparavant par l'économiste ($S_i = a + bP_i + cX_i + \gamma_i$). Ceci peut cependant, ne pas vouloir dire que les résultats seront complètement faux. Le professeur d'économétrie le clarifie avec une équation explicite de P à savoir :

$$P_i = d + eZ_i + V_i$$

où Z représente plusieurs variables qui comprennent toutes les " variables auxiliaires de pauvreté " de ciblage du programme PROSCOL. Il y aura certainement le terme d'erreur purement aléatoire qui influence la participation ; ce sont les variables auxiliaires de pauvreté qui ne sont pas dans les données et il y aura aussi des erreurs dans la sélection des participants qui figurent dans le terme V . Cette équation est linéaire, cependant P peut seulement prendre deux valeurs possibles, 0 et 1. Les valeurs comprises entre zéro et un sont acceptables mais un modèle linéaire ne peut pas exclure

la possibilité des valeurs prévus négatives, ou des valeurs au-dessus de 1. Il existe des modèles non linéaires qui peuvent traiter ce problème mais pour simplifier le débat, le plus facile sera de se concentrer sur les modèles linéaires.

Il y a un cas spécial dans lequel la régression des MCO susmentionnées de S sur P et X donnera une estimation exacte de b. C'est quand X comprend toutes les variables dans Z qui influence la sélection et le terme d'erreur V est non corrélé avec le terme d'erreur ? dans la régression de la scolarisation. Cela est parfois appelé "le choix sur des observables" dans l'exposé d'évaluation.

Supposons que les variables de contrôle X dans la régression précédente pour la scolarisation comprennent toutes les variables observées Z qui influencent la participation P et que V est non corrélé avec ? (de telle sorte que les variables non observées affectant le P du programme n'influencent pas la scolarisation conditionnée par X) ; cela a alors éliminé n'importe quelle possibilité que P soit corrélé avec ?. Il sera maintenant exogène dans la régression pour la scolarisation. Autrement dit, l'idée principale de la sélection des variables observables est qu'il y a un observable X tel que la distorsion annule la conditionnalité sur X.

L'addition des variables de contrôle à la régression de la scolarisation sur la participation au programme PROSCOL a fait une grande différence parce que X doit comprendre les variables qui étaient parmi les variables auxiliaires de la pauvreté pour le ciblage ou étaient corrélées avec elles et sont des variables qui ont aussi une influence sur la scolarisation. Cela, cependant, marche seulement si les hypothèses sont valables. Il faudra être conscient qu'il y a deux problèmes. D'abord la méthode sus mentionnée s'écroule s'il n'y a pas de déterminants de participation non observée ; autrement dit, si le terme d'erreur v a une variance 0 et que tous les déterminants de participation affectent aussi la scolarisation. Alors il n'y a aucune variation indépendante dans la participation au programme pour permettre d'identifier son impact sur la scolarisation ; il est possible de prévoir parfaitement P à partir de X et donc la régression n'estimera pas. Ce problème ne va pas surgir probablement souvent, étant donné qu'il y a presque toujours des déterminants non observés du placement du programme.

Le deuxième problème est plus courant et plus préoccupant dans ce cas. Le terme d'erreur ? dans la régression de la scolarisation contient probablement des variables qui n'ont pas été trouvées dans l'enquête, mais pourraient bien influencer la participation au programme, c'est-à-dire qu'elles pourraient être corrélées avec le terme d'erreur i dans l'équation de la participation. Si c'est le cas, alors $E(\epsilon_i | X, P) \neq 0$, et des méthodes de régression ordinaires seront toujours biaisées quand les régressions de la scolarisation seront estimées. Ainsi le problème clef est la mesure de la corrélation entre le terme d'erreur dans l'équation pour la participation et celui dans l'équation pour la scolarisation.

Exploration de meilleures voies pour former un groupe de comparaison – harmonisation du score de Propension

Avec de nouvelles informations du professeur, l'analyste apprend qu'il y a de meilleures façons de former un groupe de comparaison. L'objectif est de comparer

des niveaux de scolarisation conditionnels par les caractéristiques observées. Si les groupes types sont répartis en groupes de familles avec les mêmes valeurs ou des valeurs semblables de X , on compare les groupes conditionnels pour des familles PROSCOL et non PROSCOL. Si la scolarisation en l'absence du programme est indépendante de la participation, étant donné X , alors la comparaison donnera une estimation non biaisée de l'impact du PROSCOL. Cela est parfois appelé "l'indépendance conditionnelle" et c'est l'hypothèse clé faite par toutes les méthodes de groupe de comparaison.

Ainsi, une meilleure façon de choisir un groupe de comparaison, compte tenu les données existantes, est d'employer comme contrôle pour chaque participant un non participant avec les mêmes caractéristiques observées. Cela pourrait, cependant, être très difficile parce que le jeu des données pourrait présenter beaucoup de ces variables. Il ne peut y avoir personne parmi les non participants présentant exactement les mêmes valeurs de toutes les caractéristiques observées que n'importe quel participant PROSCOL.

Une approche statistique, l'harmonisation du score de propension, fournit des techniques pour simplifier énormément le problème. Au lieu de chercher à s'assurer que le contrôle harmonisé pour chaque participant ait exactement la même valeur X , le même résultat peut être atteint en harmonisant la valeur prévue de P compte tenu de X qui est appelé le score de propension de X . Rosenbaun et Rubin (1983) démontrent que si (dans ce cas) la scolarisation sans PROSCOL est indépendante de la participation étant donné X donc les participants sont aussi indépendants de la participation compte tenu du score de propension de X . Puisque le score de propension est juste un nombre, il est beaucoup plus facile de le contrôler que X , qui pourrait avoir beaucoup de variables. Et cependant le score de propension harmonisé est suffisant pour éliminer la distorsion puisqu'il y a l'indépendance conditionnelle compte tenu de X .

En d'autres termes, on régresse d'abord P sur X pour obtenir la valeur prévue de P pour chaque valeur possible de X , qui est alors estimé pour l'échantillon entier. Pour chaque participant, il faut trouver le non participant avec la valeur la plus proche de cette probabilité prévue. La différence dans la scolarisation est alors le gain estimé du programme pour ce participant.

On peut alors prendre la moyenne de toutes ces différences pour estimer l'impact ou bien prendre la moyenne pour les différents groupes de revenus. Cependant, cela, nécessite des précautions dans la façon dont le modèle de participation est estimé. Un modèle linéaire pourrait donner des probabilités prévues irrégulières au-dessus de 1 ou négatives. Il vaut mieux utiliser la commande de régulation LOGIT du paquet statistique. Cela suppose que le terme d'erreur v dans l'équation de la participation a une distribution logique et on estime les paramètres compatibles avec cette hypothèse par des méthodes de probabilité maximale. Ceci est basé sur les principes de l'estimation de probabilité maximale de modèles de réponse binaire.

Une autre question dont il faut être conscient est que certains des non participants peuvent avoir été exclus comme correspondants potentiels dès le départ. En fait, il y a quelques résultats récents dans le traité de l'économétrie indiquant que l'échec de la comparaison des participants et des contrôles, à des valeurs communes de variables d'harmonisation,

est une source majeure de distorsion dans les évaluations (voir Heckman et d'autres, 1998).

L'intuition est que l'on veut que le groupe de comparaison soit aussi semblable que possible au groupe de traitement en termes observables, comme récapitulé par le score de propension. Nous pourrions constater que certains échantillons de non participants ont un score de propension inférieur à n'importe lequel d'entre ceux de l'échantillon de traitement. Il est parfois appelé "le manque d'appui commun". Dans la formation du groupe de comparaison, il faut éliminer ces observations du jeu de non participants pour s'assurer que seulement les gains de la même gamme de scores de propension sont comparés. Il faut aussi exclure ces non participants dont la probabilité de participation est zéro. Il est recommandé de retirer une petite proportion de l'échantillon, par exemple deux pour cent, en haut et en bas de la distribution des non participants en terme de scores de propension. Une fois que les participants ont été identifiés et les non participants aussi sur une région d'harmonisation commune, il est recommandé de prendre une moyenne, disons environ 5 ou les nombres les plus proches, en terme de différence absolue dans les scores de propension (Encadré 3.1).

Encadré 3.1 Etapes de l'harmonisation du score de propension

Le but de l'harmonisation consiste à trouver le groupe de comparaison le plus proche d'un échantillon de non participants à l'échantillon de participants au programme.

"Le plus proche" est mesuré en termes de caractéristiques observables. S'il y a seulement une ou deux de ces caractéristiques, alors l'harmonisation doit être facile. Mais pratiquement, il y a beaucoup de caractéristiques potentielles. Les principales étapes dans l'harmonisation basée sur le plus grand nombre de propensions sont les suivantes :

Etape 1 : Vous avez besoin d'une enquête par sondage représentative pour les non participants éligibles aussi bien que pour les participants. Plus grand est l'échantillon de non participants éligibles, mieux s'en faut pour faciliter une bonne harmonisation. Si les deux échantillons proviennent d'enquêtes différentes, elles doivent être des enquêtes hautement comparables (le même questionnaire, les mêmes intervieweurs ou la même formation d'intervieweur, la même période d'enquête, etc).

Etape 2 : Regroupez les deux échantillons et estimez un modèle de régulation LOGIT de la participation au programme comme une fonction de toutes les variables dans les données qui vont probablement déterminer la participation.

Etape 3 : Créez les valeurs prévues de la probabilité de la participation à partir de la régression de la régulation LOGIT ; celles-ci sont appelées "les scores de propension". Vous aurez un score de propension pour chaque participant et non participant de l'échantillon.

(L'encadré continue sur la page suivante)

Encadré 2.2 (suite)

Etape 4 : Certains de l'échantillon de non participants devraient être exclus du jeu au début parce qu'ils ont un score de propension qui est à l'extérieur de la gamme (typiquement trop bas) observée pour l'échantillon de traitement. La gamme de scores de propension estimée pour le groupe de traitement devrait correspondre de près à celle du sous échantillon retenu de non participants. Vous pouvez aussi vouloir limiter des harmonisations potentielles d'autres façons selon le contexte. Par exemple, vous pouvez vouloir permettre seulement des harmonisations dans le même secteur géographique pour s'assurer que les parties d'harmonisation viennent du même environnement économique.

Etape 5 : Pour chaque individu dans l'échantillon de traitement, vous voulez maintenant trouver l'observation dans l'échantillon de non participants qui a le score de propension le plus proche, comme mesuré, par la différence absolue dans les scores. Cela est appelé "le voisin le plus proche". Vous pouvez trouver disons les 5 voisins les plus proches.

Etape 6 : Calculez la valeur moyenne de l'indicateur de résultat (ou chacun des indicateurs) s'il y a eu plus d'un pour les cinq voisins les plus proches. La différence entre cette moyenne et la valeur réelle pour l'observation traitée est l'estimation du gain en raison du programme pour cette observation.

Etape 7 : Calculez la moyenne de ces bénéfices individuels pour obtenir le gain moyen général. Cela peut être stratifié par quelque variable d'intérêt, comme le revenu, dans l'échantillon de non participants.

C'est la forme la plus simple de l'harmonisation du score de propension. Des complications peuvent surgir en pratique. Par exemple, s'il y a un surnombre de l'échantillonnage de participants, vous pouvez employer des méthodes de prélèvement des échantillons à base de sélection pour le corriger (Manski et Lerman 1977) ; alternativement vous pouvez utiliser le ratio de chance ($p/(1 - p)$ où p est le score de propension) pour l'harmonisation. Au lieu de compter sur le voisin le plus proche, vous pouvez plutôt employer tous les participants comme des parties potentielles mais il faut les pondérer différemment selon leur degré de proximité (Heckman et d'autres, 1998).

Ensuite, toutes les variables dans le jeu des données qui sont ou pourraient être des variables auxiliaires pour les indicateurs de la pauvreté qui ont été utilisées dans la sélection des participants au programme PROSCOL devraient être incluses. De nouveau, X doit comprendre les variables dans Z . Cela, cependant, révèle une faiblesse de l'harmonisation du score de propension. Avec l'harmonisation, X différent fournira une évaluation différente d'impact. Avec la sélection aléatoire, l'expérience idéale, les résultats ne dépendent pas de quel X vous choisissez. La randomisation ne nécessite pas non plus que l'on spécifie un modèle pour la participation ou d'une mesure de régulation LOGIT ou autre. L'encadré 3.1 récapitule les étapes à suivre pour procéder à l'harmonisation du score de propension

Apprendre les distorsions dues aux variables inobservables

Même après la formation du groupe de comparaison, l'analyste ne peut pas être sûr d'obtenir une meilleure évaluation de l'impact des programmes. Les méthodes décrites ci-dessus élimineront seulement la distorsion s'il y a l'indépendance conditionnelle, telle que les déterminants inobservables de la scolarisation non-compris dans le champ des variables de contrôle X ne sont pas corrélés avec le placement du programme. Il y a deux sources distinctes de distorsion, celles dues aux différences dans les variables observables et celles dues aux différences dans les variables non observables la dernière est souvent appelée "distorsion de sélection». L'encadré 3.2 traite de cette différence.

Un rappel de la dernière équation du professeur montre que l'indépendance conditionnelle tiendra si P est exogène, donc $E(\epsilon_i | X_i, P_i) = 0$. Cependant, le placement du programme endogène en raison du ciblage intentionnel basé sur des variables non observables laissera toujours une distorsion. Cela est parfois appelé la sélection des observables. Ainsi les conditions nécessaires pour justifier la méthode examinée auparavant par l'économiste ne sont pas moins restrictives que celles nécessaires pour justifier une version de la première méthode basée sur la comparaison des familles PROSCOL avec des familles non PROSCOL pour des ménages avec des valeurs semblables de X. Toutes les deux reposent sur la croyance que ces variables non observables n'influencent pas conjointement la scolarisation et la participation au programme, conditionnée par X.

Intuitivement, on pourrait penser qu'une harmonisation soignée réduit la distorsion mais il n'en va pas nécessairement ainsi. L'harmonisation élimine une partie de la distorsion dans la première évaluation brute de l'impact du PROSCOL. Cela laisse la distorsion due à certaines ennuyeuses variables non observées. Cependant, ces deux sources de distorsion pourraient se compenser - une positive, l'autre négative. Heckman et d'autres (1998) ont traité ce point. Donc l'harmonisation de l'évaluation pourrait bien avoir plus de distorsion que l'évaluation première. On ne peut pas savoir sur des bases préétablies de combien meilleure l'une est, même avec un groupe de comparaison bien choisi, ce qui est une question empirique.

Examiner ce qui pourrait avoir été fait avec une enquête de données de base – Estimation de la différence double

L'analyste cherche à savoir ensuite s'il y aurait une méthode autre que la sélection aléatoire qui peut contrecarrer ces non observables variables gênantes. Cela nécessiterait la collecte des données de base tant pour les participants que pour les non participants avant le démarrage du PROSCOL. L'idée est que les données collectées sur les résultats et leurs déterminants aussi bien avant qu'après que le programme ne fut introduit, avec des données pour un groupe de comparaison non traité aussi bien que le groupe de traitement. Il est alors possible de soustraire tout simplement la différence entre la scolarisation du groupe des participants et de groupe de comparaison avant le programme de la différence après le programme.

Encadré 3.2 Sources de distorsion dans les évaluations premières de l'impact du PROSCOL

La distorsion décrite par le statisticien est la différence attendue de la scolarisation sans PROSCOL entre des familles choisies le programme et celles qui n'ont pas été choisies. Cela peut être réparti en deux sources de distorsion :

- Distorsion en raison des différences de caractéristiques observables. Cela peut provenir de deux façons. D'abord, il ne peut pas y avoir de support commun. "Le support" est l'ensemble des valeurs des variables de contrôle pour lesquelles les résultats et la participation au programme sont observés. Si le support est différent entre l'échantillon de traitement et le groupe de comparaison alors cela biaisera les résultats. En effet, on ne compare pas des choses de même nature. Deuxièmement, même avec le support commun ; la répartition des caractéristiques observables peut être différente dans la région d'appui commun, en effet les données du groupe de comparaison sont mal pondérées. Un choix judicieux du groupe de comparaison peut éliminer cette source de distorsion.
- Distorsion due aux différences dans les variables non observables. Le terme distorsion de sélection est parfois limité seulement à cette composante (quoique certains auteurs emploient ce terme pour la distorsion totale dans une évaluation non expérimentale). Cette source de distorsion apparaît quand, pour les valeurs données de X , il y a un rapport systématique entre la participation au programme et les résultats en l'absence du programme. Autrement dit, il y a des variables non observés qui influencent conjointement la scolarisation et la participation au programme conditionnée par les variables observées dans les données.

Rien ne garantit que ces deux sources de distorsion fonctionneront dans la même direction. Ainsi éliminer l'une ou l'autre d'entre elles ne signifie pas en soi que toute la distorsion est réduite en valeur absolue. C'est une question empirique. Dans l'une des rares études qui traitent de cette question, le vrai impact, tel que mesuré par une expérience bien conçue a été comparé avec des évaluations non expérimentales variées (Heckman et d'autres 1998). La distorsion dans l'évaluation première était énorme, mais une harmonisation judicieuse du groupe de comparaison basé sur des variables observables a énormément réduit les distorsions.

On l'appelle l'estimation de la "différence double" ou la "différence dans des différences". Cela traitera les variables non observées ennuyeuses pourvu qu'elles ne varient avec le temps. On peut l'expliquer en ajoutant des indices à l'équation précédente de telle sorte que la scolarisation après l'introduction du programme :

$$S_{ia} = a + bP_i + cX_{ia} + \mu_{ia}$$

Avant le programme, dans l'enquête de base la fréquentation scolaire est plutôt

$$S_{ib} = a + cX_{ib} + \mu_{ib}$$

(Bien sûr $P = 0$ avant que le programme ne soit initié). Les termes d'erreur comprenant un effet invariant de temps supplémentaire, ainsi nous pouvons les écrire comme

$$\mu_{it} = \mu_i + \mu_{it} \text{ (avec } t = a, b)$$

où μ_i est l'effet non variant de temps, ce que l'on peut corrélérer avec P_i , et μ_{it} est une erreur d'innovation, qui n'est pas corrélée avec P_i (ou X_i).

L'idée essentielle ici est d'employer les données de base pour révéler ces problématiques variables non observables. Remarquez que lorsque l'enquête de base est pour les mêmes ménages que ceux que nous avons maintenant, le ménage i dans l'équation de S_a est le même dans l'équation de S_b . Nous pouvons alors prendre la différence entre l'équation "après" et l'équation "avant".

$$S_{ia} - S_{ib} = bP_i + c(X_{ia} - X_{ib}) + \mu_a - \mu_b.$$

Il est maintenant possible de régresser le changement dans la scolarisation sur la participation au programme et des changements de X . Les OMC vous donneront une évaluation exacte de l'impact du programme. Les variables non observables - celles avec la participation au programme - ont été éliminées.

Etant donné que si le placement du programme était basé seulement sur des variables tant observées que non observées qui étaient connues au moment de l'enquête, alors il serait raisonnable de supposer que les μ 's ne changent pas entre les deux enquêtes. Cela nécessiterait une bonne connaissance du programme aussi longtemps que les variables non observables problématiques sont invariantes. Les changements dans la scolarisation au cours du temps pour le groupe de comparaison révéleront ce qui serait arrivé au groupe de traitement sans le programme.

Cela nécessiterait de bien connaître le programme et la capacité de chronométrer les temps des enquêtes d'évaluation afin de les coordonner avec le programme. Autrement il subsistera vraisemblablement des variables non observés des changements après l'enquête de base qui influence celui qui

obtient un programme. Cela créerait 0's qui a changé entre les deux enquêtes.

Cette dernière équation peut être interprétée comme la signification que l'enfant et les caractéristiques des ménages dans X sont sans rapport avec le changement dans la scolarisation, si ces caractéristiques ne changent pas dans le temps. Mais le gain dans la scolarisation peut dépendre de l'éducation des parents (et non juste quelque changement dans leur éducation), et probablement de l'endroit où vit le ménage, parce que cela déterminerait l'accès aux écoles. Il peut aussi y avoir des situations dans lesquelles les changements dans le temps de l'indicateur de résultat sont sous l'influence des conditions initiales. Alors on voudra aussi contrôler pour des différences dans les conditions initiales. Cela peut se faire en ajoutant simplement X_a et X_b dans la régression séparément pour que la régression prenne la forme :

$$S_{ia} - S_{ib} = bP_i + c_a X_{ia} + c_b X_{ib} + \mu_a - \mu_{ib}.$$

Même si certaines (toutes les) variables dans X ne varient pas dans le temps on peut toujours permettre à X d'affecter les changements dans le temps de la scolarisation.

Le score de propension correspondant à la méthode traitée ci-dessus peut aider à assurer que le groupe de comparaison est semblable au groupe de traitement avant de faire la double différence. Dans une étude intéressante d'un programme d'emploi américain, il a été constaté que la défaillance à assurer que les comparaisons avaient été faites dans une région d'appui communautaire était une source principale de distorsion dans l'évaluation de la double différence en comparaison d'un groupe témoin aléatoirement sélectionné. Dans la région d'appui communautaire, cependant, la distorsion conditionnée par X n'a pas varié beaucoup dans le temps. Ainsi, il a du sens de prendre la double différence après que l'harmonisation a été faite (voir Heckman et d'autres 1998).

Cependant, en pratique, le suivi des ménages dans des enquêtes peut être difficile. Il peut ne pas être facile de trouver tous les ménages qui ont été à l'origine inclus dans l'enquête de base. Certaines personnes qui ont participé à l'enquête de base peuvent ne pas vouloir être interviewées de nouveau ou elles peuvent s'être déplacées vers un endroit inconnu.

Si les abandons au sein de l'échantillon sont purement aléatoires, l'enquête complémentaire sera toujours représentative de la même population que dans l'enquête de base. Cependant, il y a problème lorsqu'il y a quelque tendance systématique pour des gens ayant certaines caractéristiques à abandonner l'échantillon. Ceci est appelé la "distorsion d'usure". Par exemple PROSCOL pourrait aider quelques familles pauvres à avoir un meilleur logement. Et même, quand le choix du participant était seulement basé sur l'information disponible à ou autour de la date de départ des données de base (effet invariant du temps 0_i), les participants sélectionnés peuvent bien partir volontairement sur la base de changements après cette date. Une telle usure du groupe de traitement biaisera nettement une évaluation d'une différence double de l'impact du programme. L'encadré 3.3 souligne les étapes pour réaliser une double différence.

Encadré 3.3 Réaliser une différence double

La méthode de différence double entraîne la comparaison d'un groupe de traitement avec un groupe de comparaison (comme cela pourrait idéalement être déterminé par la méthode de l'harmonisation dans l'encadré 3.2) tant avant qu'après l'intervention. Les principales étapes sont les suivantes :

- Etape 1 : Vous avez besoin d'une enquête de base avant que l'intervention ne soit mise en place et l'enquête doit couvrir tant des non participants que des participants. Si vous ne savez pas qui participera à l'enquête, vous devez faire une supposition avisée. Parlez aux administrateurs du programme.
- Etape 2 : Vous avez alors besoin d'une ou plusieurs enquêtes complémentaires après la mise en place du programme. Celle-ci doit être très comparable aux enquêtes de base (en termes de questionnaire, d'interview et ainsi de suite). Idéalement, les enquêtes complémentaires doivent avoir les mêmes observations d'échantillons que l'enquête de base. Si ce n'est pas possible, alors elles doivent être des mêmes groupes géographiques ou strates en termes de certaines autres variables.
- Etapes 3 : Calculez la différence moyenne entre les valeurs antérieure et postérieure de l'indicateur du résultat pour chacun des groupes de comparaison et de traitement.
- Etape 4 : Calculez la différence entre ces deux différences moyennes. C'est votre estimation de l'impact du programme.

C'est la version la plus simple de la différence double. Vous pouvez aussi faire un contrôle des différences de conditions initiales exogènes ou des changements de variables exogènes, permettant probablement des effets d'interaction avec le programme (pour que le gain de l'intervention soit fonction des variables observables). Un modèle de régression approprié peut permettre ces variations.

Utilisation des variables instrumentales

Etant donné qu'il n'y a aucune enquête de base sur les mêmes ménages pour faire la méthode de différence double, le professeur recommande une autre méthodologie pour obtenir une évaluation consistante pour les variables non observables gênantes - une « variable instrumentale ».

Une variable instrumentale est la solution classique pour le problème de régression endogène. Une variable instrumentale est une source observable de variation exogène dans la participation au programme. Autrement dit, elle est

corrélée avec P mais n'est pas déjà dans la régression pour la scolarisation et n'est pas corrélée avec le terme d'erreur dans l'équation de scolarisation e . Donc il faut avoir au moins une variable dans Z qui n'est pas dans X et n'est pas corrélée avec e . Alors l'estimation des variables instrumentales de l'impact du programme est obtenue en remplaçant P avec sa valeur prévue conditionnée par Z .

Parce que cette valeur prévue dépend seulement de Z (qui est exogène) et que Z est non corrélé avec e il est maintenant raisonnable d'appliquer les OMC à cette nouvelle régression. Puisque les valeurs prévues dépendent seulement de la variation exogène en raison de la variable instrumentale et des autres variables exogènes, les variables non observables ne sont pas ennuyeuses parce qu'elles ne seront pas corrélées avec le terme d'erreur dans la régression de la scolarisation. Cela suggère aussi une autre façon, plus efficace, de traiter le problème. Il faut se rappeler que la source de distorsion dans l'évaluation précédente de l'impact du programme était la corrélation entre le terme d'erreur dans l'équation de scolarisation et celui dans l'équation de participation. C'est ce qui crée la corrélation entre la participation et le terme d'erreur dans l'équation de scolarisation. Ainsi une façon naturelle de se débarrasser du problème quand on a une variable instrumentale est d'ajouter les résidus de l'équation de l'étape première de la participation à l'équation de la scolarisation mais en gardant la participation réelle dans la régression de la scolarisation. Cependant, puisque nous avons maintenant ajouté à la régression de la scolarisation la valeur estimée du terme d'erreur de l'équation de la participation, il est possible de traiter la participation comme exogène et diriger les OMC. Ceci marche seulement s'il y a un instrument valable. Sinon la régression ne permettra pas d'estimer parce que la participation résiduelle sera parfaitement prévisible à partir de la participation réelle et de X dans un modèle linéaire.

Une variable instrumentale peut aussi aider s'il y a une erreur de mesure appréciable dans les données de la participation au programme, une autre source possible de la distorsion. L'erreur de mesure signifie qu'il y a la possibilité que la participation au programme varie plus qu'elle ne l'est en réalité. Cette surestimation dans P mène naturellement à une sous-estimation de son coefficient b . Ceci est appelé atténuation de distorsion parce que cette distorsion atténue le coefficient de régression.

Bien qu'une variable instrumentale puisse être extrêmement utile, en pratique il est nécessaire d'être prudent. Quand la participation réelle est tout juste remplacée par sa valeur prévue et que les OMC sont menés, cela n'indiquera pas les vraies erreurs standard parce que l'ordinateur ne saura pas que des paramètres précédemment estimés devraient être utilisés pour obtenir les valeurs prévues. Une correction des erreurs standard des OMC est nécessaire quoi qu'il y ait les outils statistiques qui permettent de le faire facilement, du moins pour des modèles linéaires.

S'il y avait une variable dépendante, cependant, qui pourrait seulement prendre une valeur possible, à l'école ou pas à l'école par exemple, alors on devrait employer un modèle de réponse binaire non linéaire, comme le modèle probit ou de régulation LOGIT. Le principe

de contrôle de l'exogénéité de la participation au programme est semblable dans ce cas. Il y a une documentation par Rivers et Vuong (1988) qui traite le problème pour de tels modèles ; Blundell et Smith (1993) fournissent une vue d'ensemble utile de divers modèles non linéaires dans lesquels il y a un régresseur endogène.

Tester les méthodologies

Quand l'analyste commence à réfléchir à l'identification d'une variable instrumentale, elle se rend compte que ce n'est pas un processus facile. Chaque possibilité qu'elle a inventée pourrait aussi être associée aux variables dans X. Le problème consiste à trouver une « restriction d'exclusion » valable qui justifie l'inclusion de quelques variables dans l'équation de la participation mais pas dans celle de la scolarisation.

L'analyste décide d'essayer la méthode de l'harmonisation du score de propension. Le modèle LOGIT de la participation semble tout à fait sensible et suggère que PROSCOL soit bien ciblé. Pratiquement toutes les variables que l'on s'attendait à voir associées à la pauvreté ont des coefficients positifs et significatifs. L'analyste réalise alors l'harmonisation du score de propension. Dans une comparaison des taux moyens d'inscription scolaire harmonisés, les résultats montrent que les enfants du groupe de la comparaison-harmonisée avait un taux d'inscription de 60 pour cent par rapport à 80 pour cent pour des familles PROSCOL.

Tenant compte de la question de revenu prévu, l'analyste fait appel à une enquête existante sur le travail d'enfant qui a interrogé sur leurs revenus. (Dans ce pays en développement, il y a une interdiction officielle du travail des enfants avant l'âge de 16 ans, mais le gouvernement arrive difficilement à mettre en application cette interdiction ; néanmoins les revenus du travail des enfants sont une question sensible). A partir de cette enquête, les gains qu'un enfant aurait faits si il ou elle n'était pas parti à l'école peuvent être déterminés.

Il est alors possible de soustraire du paiement comptant du PROSCOL aux participants la somme du revenu prévu et mettre ainsi au point le transfert du revenu net. En soustrayant ce transfert net du revenu total, il est possible de localiser d'où viennent les participants PROSCOL à la répartition du revenu pré-intervention. Ils ne sont pas tout à fait aussi pauvres que nous les avons imaginés de prime abord (ignorant le revenu) mais ils sont toujours pauvres ; par exemple, les deux tiers d'entre eux vivent en dessous du seuil de pauvreté officiel du pays.

Ayant calculé le revenu net de tous les participants, il est maintenant possible de calculer le taux de pauvreté avec et sans PROSCOL. Le taux de pauvreté après intervention (avec le programme) est simplement exposé, la proportion de la population vivant dans des ménages avec un revenu par personne en dessous du seuil de pauvreté, où le « revenu » est le revenu observé (incluant les recettes) des transferts bruts de PROSCOL. Ceci peut être calculé directement à partir de l'enquête sur les ménages. En soustrayant le gain de revenu net des transferts monétaires de PROSCOL moins le revenu attendu du travail des enfants attribué à PROSCOL

de tous les revenus observés, les résultats indiquent une nouvelle répartition des revenus avant intervention. Le taux de pauvreté sans le programme est alors la proportion des gens vivant dans des ménages pauvres basés sur cette répartition. L'analyste constate que le taux de pauvreté observé dans le Nord-Ouest de 32 pour cent aurait été 36 pour cent si PROSCOL n'avait pas existé. Le programme permet à 4 pour cent de la population d'échapper maintenant à la pauvreté. Les gains en scolarisation signifient qu'il y aura aussi des gains tant pécuniaires que non pécuniaires chez le pauvre à l'avenir. Dans le processus de mesure de la pauvreté, l'analyste se rappelle la leçon que la proportion des gens en dessous du seuil de pauvreté est seulement une mesure de base parce qu'elle ne nous dit rien des changements au dessous de la ligne (voir encadré 3.4). Elle calcule ensuite aussi bien l'indice d'écart de pauvreté que l'indice d'écart de pauvreté relevé et les résultats suggèrent que celui-ci a aussi chuté suite à PROSCOL.

Encadré 3.4 Mesures de la pauvreté

La mesure de pauvreté la plus simple et la plus fréquente est l'indice de recensement. Dans ce cas, c'est la proportion de la population vivant dans des ménages avec le revenu par personne en dessous du seuil de pauvreté. (Dans d'autres pays, c'est une mesure à base de consommation, qui présente quelques avantages ; pour le traité et les références, voir Ravallion 1994).

L'indice du recensement ne nous dit rien de la répartition des revenus en dessous du seuil de pauvreté : une personne pauvre peut être dans une situation pire mais l'indice du recensement ne changera pas, il ne reflètera pas non plus des gains parmi les pauvres à moins qu'ils ne franchissent le seuil de pauvreté.

Une alternative très utilisée de l'indice du recensement est l'indice d'écart de pauvreté (IEP). L'écart de pauvreté pour chaque ménage est la différence entre le seuil de pauvreté et le revenu du ménage ; pour ceux au-dessus du seuil de pauvreté, l'écart est de zéro. Quand l'écart de pauvreté est normalisé par le seuil de pauvreté et on calcule sa moyenne pour tous les ménages (pauvre ou non), on obtient l'indice de l'écart de pauvreté.

L'indice d'écart de pauvreté donne le degré de l'impact que le programme a eu sur la gravité de la pauvreté mais il ne reflètera pas les changements de la répartition parmi les pauvres dus au programme. Par exemple, si le programme entraîne un petit gain à une personne pauvre qui est au-dessus du revenu moyen du pauvre, en compensation d'une perte égale pour quelqu'un au-dessous de cette moyenne, alors l'indice d'écart de pauvreté ne changera pas.

(L'encadré continue sur la page suivante)

Encadré 3.4 (suite)

Il y a diverses mesures à « distribution-sensible » qui reflèteront de tels changements de la distribution parmi les pauvres. Une telle mesure est « l'écart de pauvreté relevé » (Foster, Greer, et Thorbecke 1984). Ceci est calculé de la même manière que l'indice d'écart de pauvreté sauf que les écarts de la pauvreté individuels comme proportion du seuil de pauvreté sont relevés avant de prendre la moyenne (de nouveau tant sur les pauvres que les non-pauvres). Un autre exemple d'une mesure de pauvreté à distribution sensible est l'indice de Watts. C'est la moyenne circulaire du ratio du seuil de pauvreté par rapport au revenu, où ce ratio est fixé à 1 pour le non-pauvre. Atkinson (1987) décrit d'autres exemples dans l'exposé.

Dans ce calcul, l'analyste reconnaît aussi qu'il y a quelques incertitudes relatives au seuil de pauvreté du pays. Pour tester les résultats, elle répète le calcul sur une grande gamme de seuils de pauvreté, constatant que sur un seuil de pauvreté pour lequel 50 pour cent de la population sont pauvres quand on se base sur les revenus observés après intervention, la proportion aurait été de 52 pour cent sans PROSCOL. Sur un seuil de pauvreté que 15 pour cent n'arrivent pas à atteindre avec le programme, la proportion aurait été de 19 pour cent sans cela. En répétant ces calculs sur la gamme entière des revenus, l'ensemble « des courbes d'incidence de la pauvreté » a été tracé avec et sans le programme. Il est aussi appelé « la fonction de la distribution cumulative » (voir encadré 3.5).

Encadré 3.5 Comparaison de la pauvreté avec ou sans le programme

En utilisant les méthodes décrites dans le texte principal et les encadrés précédents, on obtient une estimation du gain pour chaque ménage. Dans les évaluations les plus simples, cela constitue juste un chiffre. Mais il est mieux de le laisser varier avec les caractéristiques du ménage. On peut alors récapituler cette information en forme de courbes d'incidence de la pauvreté (CIP), avec et sans le programme.

- **Etape 1** : Le revenu après intervention (ou un autre indicateur du bien être) pour chaque ménage dans l'échantillon global (comprenant tant des participants que des non-participants) doit déjà exister ; ce sont des données. Vous savez aussi le nombre de personnes qui sont dans chaque ménage. Et bien sûr,

vous connaissez le nombre total des personnes dans l'échantillon (N ; ou cela pourrait être la taille de la population estimée si des taux d'échantillonnage inverses ont été utilisés pour épuiser chaque observation de l'échantillon).

- **Etape 2 :** Vous pouvez tracer cette information en forme de CIP. Cela donne (sur l'axe des ordonnées) le pourcentage de la population vivant dans des ménages avec un revenu moindre ou égal à cette valeur sur l'axe des abscisses. Pour faire ce graphique, vous pouvez commencer par le ménage le plus pauvre, marquer son revenu sur l'axe des abscisses et augmenter sur l'axe des ordonnées à 100 fois le nombre des gens dans ce ménage divisé par N . Le point suivant est la proportion des personnes vivant dans les deux ménages les plus pauvres, etc. Cela donne la CIP post intervention.
- **Etape 3 :** Calculez maintenant la distribution des revenus pré-intervention. Pour l'obtenir, il faut soustraire le gain estimé pour chaque ménage de son revenu post intervention. Vous avez alors une liste de revenus post intervention, une pour chaque ménage échantillonné. Répétez ensuite l'étape 2. Vous aurez alors la CIP de pré-intervention.

Si nous considérons un niveau de revenu donné sur l'axe des abscisses comme un seuil de pauvreté, donc la différence entre les deux CIP à ce point donne l'impact sur l'indice de recensement pour ce seuil de pauvreté (encadré 3.4). Alternativement, un regard horizontal vous donne le revenu correspondant à ce centile. Si aucun des gains n'est négatif alors la CIP post intervention doit se situer en dessous de celle de pré-intervention. La pauvreté aura baissé peu importe la ligne de pauvreté employée. En effet cela tient aussi pour une très large classe de mesure de pauvreté ; voir Atkinson (1987). Si quelques gains sont négatifs donc les CIP se croiseront. La comparaison de la pauvreté est alors ambiguë ; la réponse dépendra de quelles lignes de pauvreté et de quelles mesures de pauvreté on utilise (pour un exposé approfondi voir Ravallion 1994). Vous pourriez alors employer des restrictions a priori de la gamme de seuils de pauvreté admissibles. Par exemple, vous pouvez être confiant que le seuil de la pauvreté n'excède pas n'importe quelle valeur maximale et si l'intersection arrive au-dessus de cette valeur, la comparaison de la pauvreté est alors ambiguë. Si le point d'intersection (il peut en avoir plus d'un) est en dessous du seuil de pauvreté admissible maximal, donc une solide comparaison de la pauvreté est seule possible pour un jeu limité de mesures de la pauvreté. Pour vérifier comment le jeu doit être limité, vous pouvez calculer les courbes de profondeur de pauvreté (CPP). Celles ci sont obtenues en formant simplement la somme cumulative en haut de chaque point sur la CIP. (Donc le deuxième point sur la CPP est le premier point sur la CIP plus le deuxième point, etc.).

Encadré 3.5 (suite)

Si les CPP ne se croisent pas alors l'impact du programme sur la pauvreté est sans équivoque tant que l'on limite l'attention à l'indice de l'écart de pauvreté ou à n'importe laquelle des mesures de la pauvreté ou de la distribution sensible décrite à l'encadré 3.4. Si les CPP se croisent alors vous pouvez calculer « les courbes de sévérité de la pauvreté » avec ou sans le programme en formant les sommes cumulatives sous les CPP. Si celles-ci ne se croisent pas au-dessus de la gamme des seuils de pauvreté admissibles, donc l'impact sur n'importe laquelle des mesures de la pauvreté à distribution sensible dans l'encadré 3.4 est sans équivoque.

Incorporation des données du terrain

Dans la mise en œuvre de chaque programme, le point de vue des bénéficiaires et des administrateurs du programme peut être reflété dans les données du programme. Par exemple, dans ce cas la perception de ceux qui travaillent sur le terrain est que la majorité des familles PROSCOL est pauvre et que le programme est en effet utile. Quand l'analyste le discute avec un sociologue travaillant avec le programme, elle se rend compte de l'incertitude dans la réalité du manque à gagner et de la question de travail. Le sociologue discute le fait que sur le terrain on observe beaucoup d'enfants de familles pauvres qui travaillent aussi bien qu'ils vont à l'école et que certains des enfants plus jeunes qui ne sont pas à l'école ne semblent pas travailler. L'analyste se rend compte que cela nécessite quelques vérifications pour savoir s'il y a quelque différence entre la quantité du travail d'enfants fournie par des enfants PROSCOL et celle fournie par un groupe de comparaison harmonisé. Ces données, cependant, ne sont pas disponibles dans l'enquête sur les ménages qui qu'il soit possible de présenter les résultats avec et sans la déduction pour le revenu attendu.

Le sociologue a aussi remarqué que pour qu'une famille accède à PROSCOL, cela dépendra du secteur du conseil scolaire (SCS) où la famille vit importe peu. Tous les secteurs obtiennent une allocation PROSCOL du centre (SCS), même les SCS qui ont très peu de familles pauvres. Si on est pauvre mais vivant dans un SCS aisé, on a probablement plus de chance d'obtenir de l'aide du PROSCOL que lorsqu'on vit dans un pauvre SCS. Ce qui importe vraiment donc est la pauvreté relative aux autres dans le secteur où l'on vit – qu'importe beaucoup plus que le niveau absolu de vie.

Cette allocation influencerait la participation au PROSCOL mais on ne s'attendrait pas à ce que cela compte pour la scolarisation, qui dépendrait

plus du niveau absolu de vie, des conditions de famille et caractéristiques de l'école. Ainsi l'allocation du budget PROSCOL à travers les SCS peut être employées comme variable instrumentale pour supprimer la distorsion dans les évaluations d'impact du programme. Il y a l'information sur le secteur (SCS) auquel chaque ménage appartient dans l'enquête sur les ménages, les règles employées par le centre dans l'allocation PROSCOL des fonds à travers les SCS et combien le centre a alloué à chaque SCS. Les allocations de fonds sont basées sur le nombre d'enfants d'âge scolaire avec «un facteur d'ajustement» pour le degré de pauvreté où le secteur est sensé être. Cependant la règle est quelque peu vague.

L'analyste essaye de prendre ces points en considération et reprend la régression pour la scolarisation mais en remplaçant la participation réelle au PROSCOL par sa valeur prévue (le score de propension à partir de la régression pour la participation qui inclut maintenant l'allocation du budget au SCS. Cela aide à voir déjà autant de caractéristiques scolaires que possible dans la régression pour l'assiduité). Bien que les caractéristiques scolaires ne semblent pas importer officiellement pour la façon dont les ressources PROSCOL sont allouées, toutes caractéristiques scolaires omises qui influencent conjointement les allocations PROSCOL par le SCS et les résultats de scolarisation individuels laisseront une distorsion dans les estimations de variables instrumentales de l'analyste. Bien qu'il y ait toujours la possibilité de distorsion, avec beaucoup de variables de contrôle géographiques, cette méthode doit au moins offrir un élément de comparaison crédible pour l'harmonisation de l'estimation.

A partir des résultats, il est déterminé que l'allocation du budget au SCS a en effet un coefficient positif, significatif dans la régression LOGIT pour la participation au PROSCOL maintenant (prévue). La participation au PROSCOL est significative dans une régression pour l'inscription scolaire dans laquelle toutes les mêmes variables de la régression LOGIT sont incluses sauf l'allocation budgétaire au SCS. Le coefficient implique que le taux d'inscription est de 15 points en pourcentage plus élevé pour les participants au PROSCOL que cela n'aurait autrement été le cas. Une régression est aussi réalisée pour les années de scolarisation des garçons et des filles séparément. Aussi bien pour les garçons que pour les filles de 18 ans, les résultats indiquent qu'ils auraient abandonné l'école presque 2 ans plutôt sans l'intervention du PROSCOL. Cette régression, cependant, soulève des questions, à savoir si les bonnes erreurs standards vont être utilisées et si des modèles linéaires devaient être employés.

Planification du travail futur

Finalement, l'analyste est prête à publier les résultats des évaluations. Ils montrent que PROSCOL fonctionne tout à fait bien et en conséquence, les décideurs manifestent leur intérêt pour une extension du programme. Du processus par lequel l'analyste est passée pour l'exécution de l'évaluation, ressortent quelques observations importantes.

- L'évaluation d'impact peut être beaucoup plus difficile que prévu
- Il est possible d'aboutir à une gamme d'estimations préoccupantes selon les détails de la méthodologie employée.
- Il est bon d'employer des méthodes alternatives dans les situations fréquentes de données peu satisfaisantes, bien que chaque méthode à un piège ; et .
- On doit être éclectique au sujet des données.

En plus des leçons que l'analyste a tirées, elle a quelques recommandations clés pour le travail d'évaluation future de PROSCOL. D'abord, il serait souhaitable d'exclure de façon aléatoire quelques familles PROSCOL éligibles dans le reste du pays et de faire ensuite une enquête complémentaire tant sur les participants réels que sur ceux aléatoirement exclus de la participation. Cela donnerait une évaluation plus précise des bénéficiaires. Ce serait cependant politiquement sensible d'exclure certains. Cependant si le programme n'a pas assez de ressources pour couvrir tout le pays en une fois, et que le programme ait à choisir les premiers participants, il serait en effet préférable de faire cette sélection aléatoirement, parmi les participants potentiels. Alternativement, il serait possible de choisir les écoles ou des secteurs scolaires aléatoirement, dans la première vague. Une école ou un secteur scolaire serait une bonne variable instrumentale pour le placement du programme individuel.

Deuxièmement, si ce n'est pas faisable, il est recommandé d'effectuer une enquête de base dans les secteurs où il y aurait d'importantes concentrations de participants au PROSCOL avant le démarrage du programme au sud. Cela pourrait être fait en même temps que la prochaine étape de l'enquête nationale qui a été utilisée pour évaluer le programme PROSCOL. Il serait aussi bon d'ajouter quelques questions à l'enquête sur le travail rémunéré des enfants. Et troisièmement, il serait utile d'inclure un travail qualitatif qui aidera à formuler des hypothèses à tester et qui permettraient d'évaluer la plausibilité des hypothèses clés faites dans l'analyse quantitative.

Note

1. Se référer à Heckman, Lalonde, et Smith (1999), et Abadie, Angrist, et Imbens(1998) pour l'exposé sur les effets de traitement quartiles.

Chapitre 4

Exploiter les « Meilleures Pratiques » des évaluations d'impact *

Les chapitres précédents ont présenté les méthodes clés, les questions et les défis qui peuvent surgir dans l'évaluation d'impact d'un projet. Dans l'examen des études de cas inscrites au tableau 4.1, plusieurs exemples illustratifs émergent des approches intéressantes dans la conception, l'utilisation des données, le choix et l'application des méthodes analytiques et le renforcement des capacités nationales. Ces exemples, aussi bien qu'un exposé des dépenses relatives aux évaluations et aux questions d'économie politiques qui peuvent surgir dans la mise en œuvre des programmes sont mis en évidence ci-dessous.

Les 15 études de cas comprises dans la revue ont été sélectionnées à partir d'une gamme des évaluations effectuées par la Banque Mondiale, d'autres agences de donateurs, des institutions de recherches et des sociétés privées de consultation. Ils ont été sélectionnés comme un échantillon "de meilleure pratique" pour leur rigueur méthodologique, essayant de refléter une gamme d'exemples provenant de secteurs et de régions différents. Bien que chaque évaluation d'impact ait ses forces et ses faiblesses, les leçons tirées de ces expériences doivent aider le directeur de projet ou l'analyste de politique qui se propose de concevoir et de mettre en œuvre un projet futur.

Planification soignée et faite dès le début de la conception d'une évaluation

Une préparation adéquate dès les toutes premières étapes de l'identification du projet permettra de s'assurer que l'information juste est collectée et que les résultats peuvent être exploités pour les réajustements à mi-parcours des composantes du projet. Avec une planification précoce et soignée, il est possible d'incorporer tous les éléments qui contribuent à une évaluation d'impact rigoureuse comme les enquêtes de base avec un groupe témoin aléatoire et des données qualitatives sur les processus qui peuvent affecter l'impact.

Projet d'Alimentation et de Développement de la Petite Enfance en Ouganda. Cette évaluation, quoique n'étant pas encore mise en œuvre, fournit un excellent exemple de la planification faite tôt et soignée (voir Banque Mondiale 1998a; Garcia, Alderman et Rudqvist 1999). Le projet lui-même est centré sur le renforcement des capacités des parents et des communautés à s'occuper des enfants en leur fournissant une connaissance des meilleures pratiques de la protection infantile et en augmentant leurs

* Ce chapitre a été développé dans les détails dans un document d'études de cas sur les meilleures pratiques des évaluations d'impact en annexe 1 et des revues préparées par Gillette Hall et Julia Lane, Subbarao et autres (1999)

opportunités d'accroître leur revenu. Cette conception d'évaluation est basée sur la communauté et mise en œuvre par un réseau d'organisations non gouvernementales (ONG). La composante de l'évaluation qui a été intégrée au cycle du projet dès le premier jour est proche de l'idéal en terme de conception d'évaluation. D'abord il produit les données de base et des données d'enquêtes de suivi avec un groupe témoin aléatoire pour que l'impact du programme sur les résultats du bénéficiaire puisse être rigoureusement évalué. Deuxièmement elle augmente cette composante qualitative par un processus participatif (qualificatif) de suivi et d'évaluation (M & E). Du point de vue quantitatif, le projet a été conçu pour permettre une conception d'étude expérimentale dans laquelle les communes seront aléatoirement affectées dans le groupe de traitement et le groupe témoin. Les cartes de santé seront alors employées pour enregistrer des données sur le poids de l'enfant dans les communes de contrôle et de traitement. De plus l'enquête de base sur les ménages sera menée avant que les services ne soient fournis aux communautés, de même qu'une enquête de suivi sur les mêmes ménages deux ans plus tard. Un examen rapide de ces données est sensé informer la décision d'augmenter proportionnellement certaines composantes de l'intervention pendant l'examen à mi-parcours du projet. Une analyse plus approfondie des données à la fin du projet guidera la conception de la deuxième phase du projet.

Tableau 4.1 Résumé des évaluations d'impact de la « Meilleure Pratique »

Projet	Pays	Types de données	de Unité d'analyse	Mesures des résultats	Approche économétrique			Double différence	Variables instrumentales	Qualitative	Forces
					Randomi-sation	Harmoni-sation	Comparaisons réfléchies				
Education											
Radio Nicaragua	Nicaragua	Enquêtes base et post intervention	de Etudiants de salles de classe	et Evaluation par score de	Oui	Non	Oui	Non	Non	Non	Conception du questionnaire
Réforme de l'autonomie scolaire	Nicaragua	Evaluations et enquêtes par panel qualitatives	Etudiants, parents, professeurs, directeurs	Evaluation par score, degré de prise de décision locale	Non	Oui	Oui	Non	Non	Oui	M x qualitative et quantitative
Livres	Kenya	Enquêtes base et post intervention	de Etudiants, de salles de classe, professeurs	Evaluation par score de	Oui	Non	Oui	Oui	Oui	Non	Analyse des facteurs de confusion
Abandon scolaire	Philippines	Enquêtes base et post-intervention	de Etudiants, de salles de classe, professeurs	Evaluation par score de et état d'abandon scolaire	Oui	Non	Oui	Oui	Oui	Non	Analyse du coût-bénéfice; renforcement de capacité
Programmes du travail TRABAJO	Argentine	Enquête sur les ménages, recensement, fichiers administratifs, évaluations sociales	Travailleurs, ménages	Gain, coûts ciblés	Non	Oui	Non	Non	Oui	Oui	Utilisation Judicieuse des sources de données existantes, techniques analytiques innovatrices

(Le tableau continue sur la page suivante)

Tableau 4.1 Résumé des évaluations d'impact de la « Meilleure Pratique » (suite)

Projet	Pays	Types de données	de Unité d'analyse	Mesures des résultats	Approche économétrique					Forces	
					Randomisation	Harmonisation	Comparaisons réfléchies	Double différence	Variables instrumentales		Qualitative
PROBECAT	Mexique	Enquêtes sur la population active et les forces de travail rétrospectives	Travailleurs	Gains et revenus du travail	Non	Oui	Non	Non	Non	Non	Technique d'harmonisation
Programmes actifs de travail	République de Tchèque	Enquêtes sur les fichiers rétrospectifs	Travailleurs	Gains et revenus du travail	Non	Oui	Non	Non	Non	Non	Technique d'harmonisation
Finance Micro Finance	Bangladesh	Enquêtes post intervention + fichiers administratifs	Ménages	Consommation et éducation	Oui	Oui	Non	Oui	Non	Non	Analyse de facteurs de confusion
Crédit l'éducation	à Ghana	Enquêtes base et post intervention	de Paires de mère-enfant	de Revenu, santé, renforcement de capacité	Oui	Oui	Oui	Non	Non	Oui	Utilisation qualitative et quantitative de l'information

(Le tableau continue sur la page suivante)

Tableau 4.1 Résumé des évaluations d'impact de la « Meilleure Pratique » (suite)

Projet	Pays	Types de données	de Unité d'analyse	Mesures des résultats	Approche économétrique				Variables instrumentales	Forces	
					Randomisation	Harmonisation	Comparaisons réfléchies	Double différence			
Santé Financement	Niger	Enquêtes de base post intervention plus fichiers administratifs	de Ménage et de centre de santé	et Amortissement des coûts et accès	Non	Oui (dans des districts)	Oui	Non	Non	Non	Utilisation de données administratives
Alimenta-tion et nutrition Alimenta-tion pour l'éducation	Bangladesh	Enquêtes sur les dépenses des ménages	de Ménages	et Assiduité scolaire	Non	Non	Non	Non	Oui	Non	Utilisation imaginative pour traiter le problème de sélection relatif aux données de base
Santé, éducation et nutrition	Mexique	Enquêtes de base et de post intervention	de Ménages	Résultats sur la santé, l'éducation et la nutrition	Oui	Oui	Oui	Non inconnu	Non inconnu	Non	Conceptualisation claire; analyse des facteurs de confusion
Infrastruc-ture Fonds d'investis- sement social	Bolivie	Enquêtes de base et de suivi	de Ménages	Indicateur de santé et d'éducation	Oui	Oui	Oui	Oui	Oui	Non	Gamme des méthodologies d'évaluation appliquées

(Le tableau continue sur la page suivante)

Tableau 4.1 Résumé des évaluations de l'impact de la «Meilleure Pratique » (suite)

<i>Projet</i>	<i>Pays</i>	<i>Types de données</i>	<i>de Unité d'analyse</i>	<i>Mesures des résultats</i>	<i>Approche économétrique</i>					<i>Variables instrumentales</i>	<i>Qualitative</i>	<i>Forces</i>
					<i>Randomisation</i>	<i>Harmonisation</i>	<i>Comparaisons réfléchies</i>	<i>Double différence</i>				
Routes rurales	Viêt-Nam	Enquêtes base et suivi	de Ménages de communautés	et Indicateur de bien être aux niveaux des ménages et des communes	Non	Oui	Oui	Oui	Oui	Oui	Mesure des résultats de bien être	
Agriculture Projet national d'extension	Kenya	Données base panel	de Ménages, par fermes,	Efficacité et productivité de la ferme	Non	Non	Oui	Non	Non	Non	Résultats de la pertinence politique	

Projet de crédit et d'éducation au Ghana. L'évaluation de ce projet était très complexe avec beaucoup d'étapes intermédiaires. Le projet combine les éléments d'un programme de prêt de groupe avec l'éducation sur les éléments essentiels de santé, nutrition, planification et espacement des naissances et les capacités d'organiser une petite entreprise. L'évaluation est généralement focalisée sur le statut alimentaire des enfants, la capacité économique des femmes pour investir dans l'alimentation et les soins médicaux, la connaissance et l'adoption par les femmes de l'allaitement au sein et du sevrage. Il commence par une structure conceptuelle très claire qui est illustrée ci-dessous. Ce schéma trace clairement les ressources, les bénéficiaires intermédiaires et les résultats à long terme de sorte qu'il facilite le développement de plusieurs modèles ainsi que leur interprétation. En projetant judicieusement l'évaluation et en travaillant avec un schéma à une étape précoce, il fut possible de clarifier beaucoup de points d'une conception relativement complexe (voir l'Annexe 1.6).

Approches d'évaluation en l'absence des données de base

En pratique beaucoup d'évaluations n'ont pas de données adéquates. Les évaluations sont ajoutées après qu'il était possible de faire une enquête de base ou en l'absence de groupes de comparaison. Quelques exemples de ce cas sont les évaluations du Programme : Vivres pour éducation au Bangladesh, le programme PROBECAT au Mexique, les Programmes de travail actifs tchèques, et les évaluations de TRABAJAR en Argentine. En l'absence de données de base, les contrôles doivent être mis en place en employant les méthodes harmonisées traitées dans les chapitres précédents. Cela peut, cependant, être tout à fait difficile. La technique du score de

(Le texte continue à la page 71)

propension employée dans le programme TRABAJAR en Argentine pour former un groupe témoin axé sur des données croisées des participants et des non participants au programme fournit un bon exemple.

Le projet TRABAJAR II en Argentine. Ce projet était focalisé sur une offre d'emploi à bas salaires dans de petits micro-projets d'infrastructure sociale et économique choisie par les communautés. L'évaluation de l'impact du programme a été conçue pour déterminer si les revenus des participants au programme étaient plus élevés qu'ils ne l'auraient été si le programme n'avait pas été mis en place. Les méthodes les plus généralement employées pour évaluer le revenu des ménages sans intervention du projet n'étaient pas praticables dans le cas du programme TRABAJAR : aucune sélection aléatoire n'avait été employée pour former un groupe témoin à utiliser dans la comparaison des revenus des bénéficiaires du projet; et aucune enquête sur les données de base n'était disponible, excluant la possibilité de conduire une évaluation "avant et après".

L'évaluation du programme TRABAJAR par contre a exploité des données existantes pour former un groupe de comparaison par l'harmonisation des participants et des non participants de la population nationale au programme sur un jeu de variables

socio-économiques comme la scolarisation, le genre, le logement, les perceptions subjectives du bien-être et l'appartenance aux partis politiques et associations de voisinage, en employant une technique appelée score de propension. L'étude démontre une utilisation ingénieuse des données existantes sur les ménages - (Encuesta de Desarrollo Social (EDS) - dans la formation du groupe de comparaison associée à une enquête plus petite sur les participants au programme TRABAJAR menée spécifiquement pour les besoins de l'évaluation. L'enquête a été judicieusement conçue pour employer le même questionnaire que celui de (EDS) et les mêmes équipes d'intervieweurs et étaient menées approximativement à la même période en vue de réussir l'exercice de l'harmonisation. Cette technique était possible dans le cas du programme TRABAJAR parce qu'une enquête nationale sur les ménages a été sollicitée et les évaluateurs ont pu profiter de cette enquête pour sur échantillonner les participants au programme TRABAJAR. Les mêmes équipes d'intervieweurs ont été employées aussi bien pour l'enquête nationale que pour l'enquête du projet dans la collecte des données (voir l'Annexe 1.1)

Evaluation des programmes de marché du travail tchèque. Cette évaluation a essayé de couvrir cinq programmes actifs du travail pour (a) déterminer si les participants aux différents programmes avaient plus réussi à réintégrer le marché du travail que les non participants au programme et si cela variait à travers les sous-groupes et avec les conditions du marché du travail ; et (b) déterminer la rentabilité de chaque programme et faire des suggestions pour des améliorations. L'évaluation a employé une technique d'harmonisation parce qu'aucune donnée de base n'était collectée. Les évaluateurs ont enquêté sur les participants et ont ensuite sélectionné un échantillon aléatoire de non participants. Puisque les non participants étaient systématiquement plus âgés et moins

72 EXPLOITER LES « MEILLEURES PRATIQUES » DES EVALUATIONS D'IMPACT

instruits, les évaluateurs ont eu besoin de former un groupe de comparaison raisonnable pour chaque programme. Ceci a été fait en prenant chaque participant à son tour et en le comparant à chaque individu du groupe non participant sur la base de sept caractéristiques : l'âge, le genre, l'éducation, le nombre de mois d'emploi avant l'enregistrement, la taille de la ville, la situation de famille et le dernier type d'emploi. L'harmonisation la plus proche a été alors mise dans le groupe de comparaison. Bien que cette approche soit directe, la possibilité d'une distorsion de sélection subsiste - que le groupe de non participants diffère systématiquement du groupe de participants sur la base de variables inobservables (Annexe 1.5).

Gérer les contraintes qui affectent le développement de bons contrôles

De temps en temps la sélection aléatoire ou des contrôles expérimentaux sont possibles, mais ne sont pas politiquement faisables. Dans ce cas, la sélection aléatoire peut être effectuée en profitant de toute occasion pour piloter le projet dans certains secteurs limités. Les secteurs dans lesquels le projet sera piloté peuvent au commencement être aléatoirement sélectionnés avec des futurs secteurs potentiels du

projet comme contrôles. Au fil du temps, des communautés complémentaires peuvent être aléatoirement incluses dans le projet. Trois exemples illustrent comment manipuler une situation dans laquelle la sélection aléatoire était politiquement ou autrement non faisable. Au Viêt-Nam, un projet de transport rural sera évalué avec une information limitée et sans sélection aléatoire. Au Honduras, le Fonds d'Investissement Social fournit un exemple de la manière de former un groupe témoin dans des projets dont l'objectif principal est le contrôle, employant une comparaison harmonisée ex post basée sur un simple croisement de données. L'évaluation des projets dont l'objectif principal est le contrôle peut être particulièrement difficile étant donné qu'on ne connaît pas à l'avance quels projets ou communautés participeront au programme. Et, troisièmement, l'évaluation du Fonds d'Investissement Social Bolivien dans la Région de Chaco fournit un bon exemple de la manière d'intégrer la sélection aléatoire dans des projets qui sont guidés par le contrôle et qui permet un ciblage.

Projet de Pistes Rurales au Viêt-Nam. Ce projet se propose de réduire la pauvreté en milieux ruraux en améliorant l'accès aux communautés rurales et en les reliant aux réseaux routiers des provinces et des districts. La conception de l'évaluation d'impact est centrée sur des données d'enquête de base et de suivi collectées pour un échantillon de projet et des communautés de groupe de comparaison identifiées par des techniques de comparaison - harmonisation. Les données de base et les renseignements post intervention sur des indicateurs comme le niveau des productions agricoles, la diversification des sources de revenu, les offres d'emploi, la disponibilité des marchandises, des services et des équipements et les avoirs et la distribution seront collectées à partir d'un échantillon aléatoire de communes de projet (traitement) et de non projet (comparaison). Ces données seront employées pour comparer le changement des résultats avant et après l'intervention entre les communautés du projet et de non-projet en utilisant la " double différenciation ".

EXPLOITER LES « MEILLEURES PRATIQUES » DES EVALUATIONS D'IMPACT

73

Idéalement, les communes de traitement et les communes de comparaison doivent être équivalentes dans toutes leurs caractéristiques observées et non observées, la seule différence entre elles étant que les communes de traitement tirent profit de l'intervention et les communes de comparaison non. Puisque la désignation aléatoire des groupes de traitement et de comparaison n'avait pas été faite et que les données nécessaires pour faire des choix informés et des contrôles appropriés n'étaient pas disponibles au moment du choix type, les échantillons aléatoires des communes de projet et de non-projet ont été constitués. Spécifiquement, les communes de projet ont été sélectionnées d'une liste de toutes les communes avec des projets dans chaque province. Ensuite, les communes de comparaison ont été sélectionnées d'une liste de toutes les communes restantes sans projet, mais dans les mêmes zones que les communes de traitement. En employant des informations collectées pour l'évaluation, le score de propension correspondant aux techniques sera alors employé pour assurer que les communes sans projet seront des groupes de comparaison appropriés. Tous les contrôles avec des attributs inhabituels relatifs aux communes de traitement seront éliminés de l'échantillon (Annexe 1.15).

Fonds d'Investissement Social hondurien. Le Fonds d'Investissement Social Hondurien (FHIS) (Voir Banque Mondiale 1998b) vise à améliorer les conditions de vie des groupes sociaux marginaux en finançant des micro-projets d'infrastructure sociale et économique à petite échelle. Le FHIS est une institution dont l'objectif principal répond aux initiatives des municipalités, des ministères, des organisations non gouvernementales et des groupes communautaires en finançant des investissements dans les infrastructures, l'équipement et la formation. L'évaluation d'impact du FHIS emploie des techniques d'harmonisation de la comparaison, tirant l'échantillon du groupe de traitement aléatoirement d'une liste de communautés dans lesquelles les projets du FHIS ont été mis en exécution pendant au moins un an. Le groupe de comparaison, au contraire, a été choisi d'une liste de projets en réserve - ceux qui ont été sollicités et approuvés, mais pour lesquels l'investissement FHIS n'a pas encore eu lieu. En théorie, les communautés du groupe de comparaison sont automatiquement couplées avec des communautés de projet selon le processus d'auto sélection et les critères d'approbation de projet FHIS. Une enquête sur les ménages a été conduite tant dans les communautés du groupe de traitement que dans les communautés du groupe de comparaison, complétée par une composante d'évaluation qualitative (des groupes cibles et des interviews avec des informateurs clés) menée dans un sous-ensemble des communautés du groupe de traitement. Cette évaluation initiale est une première étape vers l'établissement d'un système M & E en cours qui sera éventuellement intégré aux opérations FHIS. En particulier, les données collectées auprès des communautés avec des projets potentiels deviendront des données de base utiles à partir desquelles de futurs changements dans les indicateurs d'impact seront suivis après que l'investissement du FHIS ait lieu.

Investissements dans l'éducation dans la région Chaco en Bolivie. Les projets d'éducation financés par le Fonds d'Investissement Social bolivien (SIF) sont destinés

74

EVALUATION DE L'IMPACT DES PROJETS DE DEVELOPPEMENT SUR LA PAUVRETE

à moderniser l'infrastructure scolaire et la formation des enseignants des écoles publiques rurales. Les retards dans la mise en œuvre du projet de la région de Chaco et des fonds limités pour la modernisation de l'école ont permis de faire une évaluation expérimentale en s'assurant aussi que les écoles les plus indigentes profitent du projet. Les écoles dans la région de Chaco ont été classées selon un indice de qualité scolaire basé sur la somme de cinq indicateurs d'infrastructure scolaires et d'équipements : lumière électrique, système d'évacuation, source d'eau potable, tables par étudiant et mètre carré d'espace par étudiant. Seules des écoles au-dessous d'une valeur limite particulière avaient droit à une intervention SIF. Parmi les équipements éligibles, les écoles les moins nanties ont été automatiquement sélectionnées pour profiter des investissements financés par SIF. Le premier groupe de priorité plus grande comprenait cent vingt écoles mais les fonds étaient disponibles pour mettre à niveau seulement moins de la moitié d'entre elles. Ainsi, des écoles éligibles dans le deuxième groupe de priorité ont été aléatoirement réparties en groupes de traitement et de comparaison, fournissant les conditions pour une conception d'évaluation expérimentale (Annexe 1.4)

Combinaison des méthodologies

Pour la plupart des évaluations, plus d'une technique est nécessaire pour réaliser des résultats solides qui traitent plusieurs questions d'évaluation. Chaque question peut nécessiter des techniques différentes, même dans une seule conception de projet. Trois exemples illustrent comment plusieurs techniques ont été combinées dans une évaluation; le Fonds Social de la Bolivie, l'évaluation du programme TRABAJAR en Argentine et la réforme scolaire au Nicaragua.

Le Fonds Social de la Bolivie. Les fonds sociaux incluent généralement plusieurs différents types de micro-projets et ainsi la conception d'une évaluation peut comporter plusieurs approches. Dans le cas du Fonds Social de la Bolivie, le modèle de la mise en œuvre du projet a dicté les méthodes d'évaluation. Dans le domaine de l'éducation, les écoles qui devaient recevoir l'intervention avaient déjà été identifiées ; donc la sélection aléatoire ne pouvait pas être employée. A la place, des méthodes d'harmonisation ont été adoptées. Dans le cas des projets de santé, des méthodes réflexives ont été employées parce que l'intervention devait être mise en œuvre dans tous les centres médico-sociaux de la région (Voir Annexe 1.4).

Utilisation d'une large combinaison de composantes d'évaluation – Programme TRABAJAR II de l'Argentine. L'évaluation du programme TRABAJAR comprend une gamme de composantes conçues pour évaluer le programme de réalisation de ses objectifs politiques. La première composante s'appuie sur les données des enquêtes sur les ménages pour évaluer les acquis de revenu des participants au programme TRABAJAR. La deuxième composante contrôle l'allocation de financement du programme (ciblage), dépistant les changements sur une période de temps comme résultat de la réforme. Cette composante produit un feedback deux fois par an

pour raffiner le ciblage du programme. Des composantes d'évaluation complémentaires comprennent une analyse coûts bénéfices des projets d'infrastructures et des évaluations sociales conçues pour fournir des réactions de la communauté à la mise en œuvre du projet. Chacune de ces composantes a été menée deux fois. Trois composantes futures sont projetées. La technique de la recherche d'harmonisation de la comparaison sera appliquée de nouveau pour évaluer l'impact de la participation au programme TRABAJAR sur l'activité du marché du travail. La qualité du projet d'infrastructure sera réévaluée, cette fois-ci pour les projets qui ont été achevés pendant au moins un an pour évaluer la durabilité, l'entretien et les taux d'utilisation. Finalement, une composante de recherche qualitative examinera les opérations et les procédures du programme en interviewant des membres du personnel des agences qui sponsorisent les projets aussi bien que les bénéficiaires du projet.

Les résultats de l'évaluation fournissent une direction claire à la réforme de la politique. La première composante d'évaluation révèle que le programme TRABAJAR

a connu un grand succès par le ciblage des pauvres – l’auto sélection des participants, en offrant des bas salaires est une stratégie qui marche en Argentine et les participants apprécient les gains de revenu comme un résultat de la participation. La deuxième composante constate que l'affection géographique du financement du programme s'est améliorée au fil du temps - le programme connaît maintenant plus de succès avec l'allocation des fonds à des secteurs pauvres ; cependant, le processus d'évaluation en cours indique que la performance varie et est constamment faible dans certaines provinces, des constats auxquels une attention politique plus poussée est actuellement consacrée. Des résultats d'évaluation décevants sur la qualité du projet d'infrastructure ont entraîné des efforts énormes de l'équipe du projet visant à améliorer la performance dans ce secteur à travers une réforme politique – insistant sur des visites accrues de site pour l'évaluation et le suivi, pénalisant les agences dont la performance est médiocre à l'achèvement du projet et renforçant le manuel d'évaluation. Et finalement, les évaluations sociales ont mis à jour le besoin d'accroître une meilleure assistance technique aux organisations non gouvernementales et aux municipalités rurales pendant la préparation et la mise en œuvre du projet aussi bien qu'une plus grande publicité et une transparence de l'information sur le programme TRABAJAR (Annexe 1.1).

Réforme de l'autonomie des écoles du Nicaragua. En 1993, le gouvernement nicaraguayen a pris des mesures décisives pour mettre en œuvre une importante initiative de décentralisation dans le secteur de l'éducation, en accordant une autonomie de gestion et budgétaire à certaines écoles primaires et secondaires. Le but des réformes est de renforcer l'apprentissage de l'étudiant : comme la gestion scolaire devient plus démocratique et participative, la gestion scolaire locale et des modèles de dépenses peuvent être consacrés à des efforts qui améliorent directement la pédagogie et renforcent l'apprentissage de l'étudiant. L'impact de cette réforme a été évalué en employant une combinaison de techniques quantitatives et qualitatives pour évaluer le résultat aussi bien que le processus de la décentralisation. Le but de cette composante qualitative est de révéler si réellement ou non la

gestion projetée et les réformes de financement ont lieu dans les écoles et pourquoi ou pourquoi pas. La composante quantitative renforce ces résultats en répondant à la question "les changements dans la gestion et le financement scolaire produisent-ils en réalité de meilleurs résultats d'étude pour les enfants ?" Les résultats qualitatifs montrent que la mise en œuvre couronnée de succès des réformes dépend en grande partie du contexte scolaire et de l'environnement (c'est-à-dire le seuil de pauvreté de la communauté), tandis que les résultats quantitatifs indiquent que le processus décisionnel accru des écoles est en fait significativement associé à l'amélioration de la performance de l'étudiant.

Des méthodologies et des sources de données différentes mais complémentaires ont été employées pour combiner les deux approches. D'une part, l'évaluation quantitative a suivi une conception quasi expérimentale dans laquelle les résultats de test d'un échantillon d'étudiants des écoles autonomes (le groupe de traitement) sont comparés avec ceux d'un échantillon correspondant des écoles publiques non autonomes et des écoles privées (le groupe de comparaison). Les données pour cette composante de l'évaluation ont été collectées d'un panel de deux enquêtes couplées école-ménage et

les résultats d'un test sur la performance des étudiants. La conception de l'évaluation qualitative, d'autre part, a consisté en une série d'interviews des informateurs clés et des discussions du groupe cible avec différent personnel, des écoles et des parents dans un sous échantillon des écoles autonomes et traditionnelles comprises dans l'étude quantitative.

L'utilisation des techniques de recherche tant qualitative que quantitative a produit une combinaison appréciable de résultats utiles politiquement pertinents. Le travail quantitatif a fourni une large vue d'ensemble, statistiquement valable des conditions et des résultats scolaires; le travail qualitatif a renforcé ces résultats avec la compréhension de la raison pour laquelle certains résultats attendus du programme de la réforme avaient été couronnés de succès tandis que d'autres avaient échoué et ont donc aidé à guider des réajustements politiques. En outre, parce que c'est plus intuitif, le travail qualitatif était plus accessible et donc intéressant pour le personnel du Ministère, qui à son tour a facilité un renforcement des capacités et une crédibilité rapide pour le processus d'évaluation dans le ministère (Annexe 1.11).

Exploitation des sources de données existantes

Des sources de données existantes comme une enquête nationale sur les ménages, le recensement, le registre du programme administratif ou des données municipales peuvent souvent fournir une ressource de valeur aux efforts d'évaluation. Exploiter des sources existantes réduit le besoin de la collecte de données coûteuse pour le seul but de l'évaluation comme illustré dans le cas de l'évaluation des pistes rurales au Viêt-Nam. En outre bien que des données existantes puissent ne pas contenir toutes les informations que l'on collecterait idéalement pour les buts de l'évaluation, des techniques d'évaluation innovatrices peuvent souvent compenser des données manquantes comme indiqué dans le Projet d'Extension Agricole au Kenya.

Projet de Routes Rurales au Viêt-Nam. Les données utilisées dans cette évaluation exploitent une combinaison efficace des sources de données nationales et locales existantes avec des enquêtes conduites spécifiquement pour les buts de l'évaluation. L'évaluation de l'enquête sur les ménages est efficacement conçue pour reproduire un nombre de questionnaires de l'Enquête sur le Niveau de Vie du Viêt-Nam pour que, exploitant l'information commune aux deux enquêtes, les techniques de régression puissent être employées pour évaluer la position de chaque ménage dans la distribution nationale du bien être.

L'évaluation repose largement sur des données du niveau de la commune collectées annuellement par les communes, couvrant les données démographiques, l'utilisation de la terre et les activités de production. Cette source de données a été renforcée par une enquête du niveau de vie de la commune conduite pour les buts de l'évaluation. Deux bases de données complémentaires ont été constituées utilisant l'information existante. Une base de données de grande envergure au niveau de la province a été établie pour aider à comprendre le choix des provinces dans le projet. Cette base de données couvre les provinces de tout le Vietnam et englobe des données sur un grand nombre de

variables socio-économiques. Finalement, une base de données au niveau du projet, pour chacun des secteurs de projet enquêtés, a été aussi constituée, en vue de contrôler l'ampleur du projet et la méthodologie de mise en œuvre de l'évaluation de l'impact du projet (Annexe 1.15).

Projet National d'Extension du Kenya (PNE). La performance du Projet National d'Extension du Kenya (PNE) a été controversée et fait partie d'un plus grand débat sur la rentabilité de l'approche de formation et de visite (T&V) des services d'extension agricole. Dans le contexte kenyan, le débat a été soulevé par, des retours en force de T&V d'une part, annoncés dans une étude (Bindlish, Evenson et Gbetibouo 1993, 1997) et d'autre part, le manque de résultats visibles convaincants y compris la piètre performance de l'agriculture kenyane ces dernières années.

Le désaccord sur la performance du PNE a persisté en attendant les résultats de cette évaluation qui a été conçue pour adopter une approche rigoureuse, empirique, à l'évaluation du développement institutionnel du programme et l'impact sur la performance agricole. L'évaluation utilise une combinaison de méthodes qualitative et quantitative pour poser des questions politiques hautement pertinentes et révèle des faiblesses sérieuses dans le programme : (a) le développement institutionnel du PNE a été limité et après 15 ans d'existence, il y a peu d'amélioration de l'efficacité de ses services, (b) la qualité et la quantité de ses prestations de service sont faibles ; et (c) les services d'extension ont seulement un petit impact positif sur l'efficacité de la ferme et aucun impact sur la productivité de la ferme.

L'évaluation est en mesure de dresser un tableau des conclusions politiques concrètes de ces résultats, dont beaucoup sont relatifs à la conception de futurs projets d'extension agricole. D'abord, l'évaluation révèle un besoin de renforcer le ciblage de T&V se concentrant sur des secteurs et des groupes dont l'impact va probablement être le plus grand. En outre, le conseil doit être soigneusement modelé pour répondre

aux besoins du paysan en tenant compte des variations des conditions technologiques et économiques locales. Réaliser avec succès ce niveau de ciblage de service demande des flux appropriés d'informations opportunes et fiables dont un sphère de Programme M&E générant un flux de feedback des bénéficiaires sur le contenu du service. Pour renforcer l'efficacité du programme, une présence plus discrète et moins intense avec une couverture plus large va probablement être plus rentable. L'approche générale du programme de la prestation des services, utilisant une méthodologie unique (des visites de la ferme) pour livrer des messages standard, limite aussi l'efficacité du programme. La réforme institutionnelle va probablement renforcer l'efficacité de la prestation des services, la décentralisation du programme de conception comprenant les mécanismes participatifs qui donnent la parole au paysan (comme le partage des coûts et les organisations de paysans) doivent devenir une partie intégrante du mécanisme de la prestation de service. Finalement, l'amortissement des coûts, même partiel, fournirait des motivations appropriées, aborderait les questions de la responsabilité et du contrôle de la qualité, rendra le service plus exigeant et sensible et fournira une certaine amélioration budgétaire (Annexe 1.8).

Coûts et Financements

Il y a sans doute beaucoup de dépenses impliquées dans l'exécution d'une évaluation d'impact, ce qui explique pourquoi certains pays sont réticents à financer de telles études. Ces coûts comprennent la collecte des données et le coût du temps du personnel composant l'équipe d'évaluation. Le financement d'une évaluation d'impact peut provenir de l'intérieur d'un projet, d'autres ressources du gouvernement, d'une bourse de recherche ou un donateur extérieur. L'information sur un échantillon des évaluations de la Banque Mondiale démontre que bien que beaucoup de pays aient assumé la majorité des dépenses d'évaluation, la réussite de la mise en œuvre d'une évaluation d'impact a nécessité des ressources extérieures substantielles au-delà de celles prévues pour le prêt ou le crédit d'un projet. Ces ressources proviennent d'une combinaison des sources suivantes : (a) un prêt de la Banque Mondiale ou un financement de crédit pour la collecte et le traitement des données, (b) le gouvernement, par les salaires payés au personnel local affecté à l'effort de l'évaluation (comme expliqué dans le tableau 4.1, ces dépenses de personnel n'ont pas été incluses dans le calcul des dépenses d'évaluation fait pour les cas passés en revue ici à cause des limitations de données) ; (c) des bourses de recherche de la Banque Mondiale et des subventions de donateurs bilatéraux qui ont financé l'assistance technique des consultants avec l'expertise spécifique nécessaire pour l'évaluation, et (d) les frais généraux de la Banque Mondiale pour régler les honoraires du personnel fourni pour guider l'évaluation d'impact et qui souvent participent activement au travail d'analyse.

Bien que peu d'évaluations d'impact documentent le coût du travail, le tableau 4.2 fournit des devis, pour un échantillon d'évaluations d'impact avec la participation de la Banque Mondiale. Ces devis ne comprennent pas cependant la valeur du temps du personnel apporté comme contribution par les pays bénéficiaires

Tableau 4.2 Résumé des coûts estimés à partir de plusieurs évaluations d'impact de la Banque mondiale

<i>Projet</i>	<i>Récapitulation de l'évaluation des coûts (%)</i>						
	<i>Evaluation des coûts estimés (\$)^a</i>	<i>Coût total % du coût projet^b</i>	<i>en Coût % du coût du prêt de la BIRD ou du crédit de l'IDA^b</i>	<i>en Voyage</i>	<i>Personnel de la Banque Mondiale</i>	<i>Consul- tants</i>	<i>Collecte de données</i>
Nicaragua	–						
Gestion basée sur l'école	sur 495 000	1,20	1,5	8,1	18,1	39,0	34,8
El Salvador	–						
Gestion basée sur l'école	sur 443 000	0,60	1,3	7,7	7,4	25,8	59,2
Colombie	–						

Programme de Bon	266 000	0,20	0,3	9,4	9,8	21,8	59,0
Honduras – Fonds social	263 000	0,23	0,9	3,0	11,5	53,2	32,3
Nicaragua – Fonds social	449000	0,30	0,8	4,9	33,0	7,8	55,7
Bolivie – Fonds social	878 000	0,50	1,4	3,4	14,6	12,9	69,1
Trinité et Tobago – Formation des jeunes	des 238 000	0,80	1,2	7,6	11,5	17,9	63,1
Moyenne	433 000	0,56	1,0	6,3	15,1	25,5	53,3

a. Ces coûts ne comprennent pas les coûts des équipes locales de contrepartie qui ne sont pas financés à partir du prêt ou du crédit. Les chiffres portés en référence au titre desquels les projets en évaluation ont été sélectionnés, non pas le financement total fourni par la Banque Mondiale et d'autres institutions.

b. Ces coûts en pourcentage du prêt ou de crédit ou du projet sont présentés seulement comme référence. Dans beaucoup de cas le financement réel de l'évaluation a été obtenu de sources extérieures au financement du projet.

c. Les estimations de coût de voyage comprennent (voyage de mission pour le personnel de la Banque Mondiale et les consultants internationaux jusqu'aux pays bénéficiaires aussi bien que le voyage des partenaires des pays bénéficiaires, particulièrement pour participer aux sessions de stratégie et aux ateliers analytiques avec les consultants internationaux et le personnel de la Banque Mondiale.

Source: archives de projet de la Banque Mondiale.

partenaires (qui peut être significative) parce que cette information n'était pas disponible. Comme repère, dans les huit cas ci-dessus, il n'était pas inhabituel d'avoir jusqu'à cinq membres du personnel affectés au travail d'évaluation pendant plusieurs années, un niveau d'effort suffisant pour augmenter substantiellement le coût de l'évaluation dans beaucoup de ces cas.

80 EVALUATION DE L'IMPACT DES PROJETS DE DEVELOPPEMENT SUR LA PAUVRETE

Le coût estimatif moyen de l'évaluation de l'impact était 433 000\$. Ceci reflète un ordre de grandeur des coûts de 263 000\$ pour l'évaluation d'un programme de formation de compétences professionnelles pour la jeunesse sans emploi à Trinité et Tobago, à 878 000\$ pour l'évaluation du Fonds d'Investissement Social Bolivien. Les dépenses relatives à l'évaluation d'impact pour les projets ci-dessous reflètent en moyenne 0,6 pour cent du coût total du projet (qui incluait parfois le financement de plusieurs donateurs), ou 1,3 pour cent du coût du prêt de la Banque Internationale pour la Reconstruction et le Développement (BIRD) ou du crédit de l'Association Internationale de Développement (IDA). Les composantes les plus coûteuses des évaluations inscrites ci-dessous étaient la collecte des données et les consultants, tant locaux qu'internationaux. Dans beaucoup de cas, les dépenses de voyage comprennent le voyage du personnel local pour rencontrer le personnel de la Banque Mondiale et des chercheurs pour des sessions de stratégie et la formation parce que le renforcement des capacités du personnel du pays bénéficiaire était un objectif clé. Les deux exemples ci-dessus relatifs aux évaluations d'impact de projets à Trinité et Tobago et en Bolivie illustrent quelque variation qui peut surgir dans les dépenses de programme.

L'évaluation du programme de formation des compétences professionnelles en Trinité et Tobago a bénéficié d'une enquête sur le revenu national et sur l'emploi pour sur échantillonner des diplômés du programme et créer un groupe de comparaison à partir d'un sous-ensemble de l'échantillon national. De plus, l'équipe d'évaluation a aidé à concevoir et utiliser des données administratives disponibles, des rapports de demandeurs de programme, ainsi des données pré-intervention étaient disponibles et aucun décompte n'était nécessaire pour localiser les diplômés du programme. La taille type de l'échantillon pour chacune des trois études de contrôle était environ 2 500 jeunes gens, comprenant tant des groupes de comparaison que de traitement. Il y avait seulement un court questionnaire à administrer dans l'enquête et on a donné le questionnaire seulement aux diplômés du programme. Trinité et Tobago est un petit pays dont les communautés sont relativement faciles d'accès par la route et l'anglais est la langue commune dans le pays et parmi les diplômés du programme.

L'évaluation du Fonds Social de la Bolivie (SIF) a mené ses propres enquêtes de base et de suivi sur les groupes de traitement et de comparaison pour évaluer des interventions en santé, éducation, eau et assainissement. Il n'y avait aucune enquête nationale disponible à partir de laquelle conduire des analyses pour effectuer le sur échantillonnage, ce qui a mis la charge entière de la collecte des données sur l'évaluation. L'échantillon des groupes de traitement et de comparaison a consisté en près de 7 000 ménages et 300 équipements interviewés aussi bien dans l'enquête de base de 1993 que dans l'enquête de suivi de 1998.

En Bolivie, les instruments de la collecte des données pour l'évaluation d'impact ont consisté en des laboratoires ambulants pour conduire des tests de qualité d'eau en laboratoire, des tests de confirmation et huit questionnaires pour des informateurs provenant des ménages et des structures. Les huit questionnaires ont consisté en deux questionnaires sur le ménage (un pour l'informateur principal et un pour les femmes en âge de procréer), un questionnaire pour la communauté, quatre différents

questionnaires du centre médico-social pour les différents types de centres médico-sociaux (allant des petites cliniques de communauté aux hôpitaux) et un questionnaire scolaire pour le directeur d'école. Pour évaluer le ciblage, l'évaluation a inclus une mesure de la pauvreté basée sur la consommation qui a nécessité la collecte de données de consommation détaillées fournies par les ménages aussi bien que les données des prix régionaux fournies aux communautés. Le travail de terrain a été mené en milieux ruraux où la majorité des projets du FIS est focalisée et comprenant un échantillon aléatoire des ménages ruraux qui étaient souvent accessibles seulement à pied ou à cheval. Finalement, les questionnaires ont dû être développés et administrés en espagnol, Quechua et Aymara.

Questions d'économie politique

Il y a plusieurs questions d'économie politique qui affectent non seulement la réalisation d'une évaluation, mais aussi sa mise en œuvre. La décision de poursuivre une évaluation est beaucoup plus contingente d'un appui politique fort. Beaucoup de

gouvernements ne voient pas l'importance d'évaluer des projets et, en conséquence, ne veulent pas y investir des ressources. En plus, il peut y avoir la réticence à permettre une évaluation indépendante qui peut trouver des résultats contraires à la politique du gouvernement, particulièrement dans des régimes autoritaires ou fermés. Des gouvernements ouverts peuvent, cependant, considérer les évaluations et la diffusion des conclusions comme une partie importante du processus démocratique.

Les évaluations sont aussi sensibles au changement politique. Trois des huit études d'impact inscrites dans le tableau 4.2 ont été annulées à cause des questions d'économie politique. Un changement de régime ou de postes clés dans un bureau du gouvernement partenaire et des changements de stratégie politique peuvent affecter non seulement l'effort d'évaluation mais plus fondamentalement la mise en œuvre du programme à évaluer. Un exemple de ce type de risque vient de l'expérience d'une équipe travaillant sur la conception et l'évaluation d'impact d'une gestion pilote basée sur l'école au Pérou, comme partie d'un projet d'éducation primaire financé par la Banque Mondiale. L'équipe composée des fonctionnaires du Ministère de l'Éducation nationale, du personnel de la Banque Mondiale et des consultants internationaux et bcaux avait travaillé pendant plus d'un an à développer des modèles de gestion basés sur l'école- pilote en projet, à établir une conception expérimentale, à concevoir des instruments d'enquête et des tests de confirmation et à collecter des données de base sur des caractéristiques scolaires et la performance de l'étudiant. Juste avant l'introduction du test pilote dans les écoles aléatoirement choisies, des représentants gouvernementaux de haut niveau ont annulé l'expérience de la gestion basée sur l'école en réaction aux retombées politiques perçues du test pilote. Une réforme semblable a été présentée plusieurs années plus tard mais sans l'avantage d'un test pilote ou d'une évaluation.

Au Vénézuéla, une évaluation d'un programme de santé maternelle et infantile et de nutrition a été reprise trois fois avec trois différents partenaires bénéficiaires de

82 EVALUATION DE L'IMPACT DES PROJETS DE DEVELOEMENT SUR LA PAUVRETE

contreparties puisque le gouvernement a déplacé la responsabilité de l'évaluation d'une agence à une autre. Chaque changement a été accompagné d'une renégociation du contrat avec la société du secteur privé qui avait été identifiée pour effectuer la collecte des données et la majorité de l'analyse pour l'évaluation. Quand la légitimité de la troisième contrepartie du gouvernement a commencé à être mise en doute, la société a annulé le contrat et l'évaluation a été abandonnée ! Ces incidents sont arrivés pendant une période d'instabilité politique caractérisée par les nombreux remaniements ministériels qui ont fini par l'écroulement du gouvernement élu servant comme contrepartie pour le projet, donc l'évaluation était presque seule à souffrir de l'instabilité politique. Néanmoins, dans les cas tant péruvien que vénézuélien, il est sage de réfléchir sur l'importance des ressources consacrées à un effort qui n'a jamais été mené à bout. Un exemple moins dramatique de l'effet du changement politique sur des stratégies d'évaluation vient du Salvador, où le succès reconnu d'une réforme présentée aux écoles rurales a incité le gouvernement à présenter une réforme de l'éducation semblable dans toutes les écoles urbaines en même temps, au lieu d'une intégration progressive aléatoire dans les écoles comme originalement programmé.

Cette décision a éliminé la possibilité d'employer une conception expérimentale et a laissé l'utilisation d'une comparaison réfléchie moins rigoureuse comme la seule option d'une conception d'évaluation viable.

Annexe 1

Etudes de cas

Annexe 1.1 : Evaluation des acquis des pauvres dans le cadre du programme : protection-travail associant aide sociale et aide à l'emploi : Le Programme TRABAJAR en Argentine

I. Introduction

Description du projet. Le Programme argentin TRABAJAR vise à réduire la pauvreté, en créant simultanément des possibilités d'emploi pour les plus démunis et en améliorant l'infrastructure sociale au sein des communautés pauvres. TRABAJAR I, est un programme-pilote, qui a été mis en place en 1996, en réponse à une grave crise économique et des taux de chômage de plus de 17%. TRABAJAR II a été lancé en 1997 en tant que version élargie et corrigée du programme-pilote, et TRABAJAR III a commencé à approuver des micro-projets en 1998. Le programme offre des salaires relativement bas afin d'attirer uniquement, (Auto-sélection) en tant que participants, des travailleurs pauvres, des désœuvrés. Les projets d'infrastructure pour lesquels les participants ont été engagés sont proposés par des organisations gouvernementales et non-gouvernementales (ONGs), lesquelles doivent couvrir les coûts hors-salaire du projet Les projets sont approuvés, au niveau régional, selon les directives du gouvernement central.

A la lumière des résultats du processus d'évaluation, le programme a subi quelques modifications des procédures relatives à la conception et au fonctionnement. TRABAJAR II, comprenait un certain nombre de réformes visant à améliorer le ciblage des projets. Le système d'allocation budgétaire du gouvernement central est désormais plus largement influencé par des considérations telles que la pauvreté provinciale, ainsi que les indicateurs de chômage, et un accent plus important est mis sur les propositions de projets de la part des zones défavorisées, dans le cadre des directives d'agrément des projets. Au niveau local, des efforts ont été consentis sous TRABAJAR II et TRABAJAR III, afin de renforcer les capacités des bureaux de provinces à aider les zones défavorisées à monter des projets et à rehausser les normes de qualité des infrastructures.

Evaluation de l'impact. L'effort d'évaluation a commencé dès la préparation du projet TRABAJAR II et se poursuit sans relâche. Le but de l'évaluation est de déterminer dans quelle mesure le programme atteint les objectifs de sa politique ainsi que d'indiquer les zones pour lesquelles celui-ci nécessite des réformes, pour maximiser son efficacité. L'évaluation consiste en une série d'études séparées qui apprécient (a) les revenus qui s'accumulent au profit des participants, (b) l'allocation des ressources du programme dans les régions (ciblage), (c) la qualité des projets d'infrastructure ayant reçu un financement, et (d) le rôle de la communauté ainsi que celui des ONGs dans les résultats du projet.

Deux des composantes de l'évaluation ressortent particulièrement, sur le plan technique, démontrant les techniques empiriques de meilleures pratiques. Tout d'abord, l'étude des revenus

illustre les techniques de meilleures pratiques, par comparaison appariée¹, ainsi que l'utilisation habile des données nationales des enquêtes sur les ménages lors de l'exercice comparatif. Deuxièmement, l'étude de résultats spécifiques introduit une nouvelle technique pour évaluer le ciblage, lorsque l'incidence des dépenses publiques locales n'est pas sous contrôle. La conception globale de l'évaluation présente également une multitude de meilleures pratiques, consistant en un mélange de composantes et de techniques de recherche – de l'analyse quantitative aux visites de sites d'ingénierie en passant par l'évaluation sociale – lesquelles fournissent une palette intégrée de résultats de façon opportune.

II. Conception de l'évaluation

L'évaluation TRABAJAR inclut une gamme de composantes qui ont été conçues pour évaluer dans quelle mesure le programme a atteint les objectifs de sa politique. La première composante se base sur les données de l'enquête sur les ménages afin d'évaluer les revenus pour les participants du programme TRABAJAR. Cette étude est une amélioration des évaluations conventionnelles de programmes d'allocation conditionnelle décentralisée, lesquelles, ordinairement, mesurent les revenus des participants, simplement en tant que salaires bruts, en procédant à une estimation des revenus nets. En ayant recours aux avancées, en termes de techniques de comparaison appariée, l'étude rend compte des revenus renoncés (revenus auxquels les participants ont renoncé en rejoignant le programme TRABAJAR), dont la résultante est une estimation moindre et plus exacte des revenus nets des participants. La deuxième composante examine l'allocation de fonds du programme (ciblage), considérant les changements diachroniques dus à la réforme. A travers l'utilisation judicieuse des données généralement disponibles (l'allocation des fonds du programme dans les provinces ainsi qu'un recensement national), la conception de cette composante présente une nouvelle méthodologie afin d'évaluer le ciblage des poches de pauvreté, lorsque aucune donnée sur l'incidence du programme n'est disponible. Cette analyse a commencé avec la première mission d'encadrement (Novembre 1997). Elle a été mise à jour deux fois.

Des composantes supplémentaires d'évaluation comprennent une analyse des coûts et bénéfices réalisés pour un sous-échantillon de projets d'infrastructure, de pair avec des évaluations conçues pour apporter un certain écho sur la mise en œuvre des projets. Chacune de ces activités a été réalisée deux fois, pour chacun des programmes TRABAJAR II et III. Trois activités d'évaluation sont planifiées. La technique de recherche de comparaison appariée sera à nouveau appliquée pour évaluer l'impact de la participation au programme TRABAJAR sur le marché du travail. La qualité des projets d'infrastructure sera ré-évaluée, cette fois-ci pour des micro-projets achevés depuis au moins une année, afin d'évaluer la durabilité, la maintenance ainsi que les taux d'utilisation. Finalement, une composante recherche qualitative analysera la réalisation et les procédures du programme en s'entretenant avec les membres du personnel au sein des agences qui parrainent les projets ainsi que les bénéficiaires du programme.

III. Collecte des données et techniques d'analyse

L'évaluation des revenus nets des participants au programme se base sur deux sources d'information : : une enquête nationale sur la qualité du niveau de vie (Encuesta de Desarrollo Social - EDS) et une enquête sur les participants au programme TRABAJAR, réalisée spécialement pour les besoins de l'évaluation.. (Le financement de l'enquête EDS a été pris en charge par un autre projet de la Banque Mondiale. Celle-ci a été conçue pour améliorer la qualité de l'information sur le bien-être des ménages en Argentine, particulièrement pour ce qui concerne l'accès aux services sociaux et programmes sociaux du gouvernement). Ces enquêtes ont été réalisées en 1977, en août (EDS) et en septembre (enquête sur les participants au programme TRABAJAR) par l'Office National des Statistiques, en utilisant le même questionnaire et les mêmes équipes d'enquêteurs. L'échantillon de l'enquête EDS couvre 85% de la population nationale, à l'exception de quelques zones rurales et très

¹ utiliser un processus non-aléatoire dans la sélection d'un groupe de contrôle qui correspond le plus exactement aux caractéristiques des bénéficiaires du programme

petites communautés. L'échantillon de l'enquête sur les participants au programme TRABAJAR est réalisé à partir d'un échantillon aléatoire du projet TRABAJAR II situé dans le cadre des échantillons EDS. Il génère des données pour 2.802 participants actuels (le total, pour TRABAJAR II, entre mai 1997 et janvier 1998 atteint 65.321 participants). La fiabilité de la technique de comparaison appariée est maximisée par la possibilité de soumettre le même questionnaire, au même moment, aux participants ainsi qu'au groupe de contrôle, et de s'assurer que les deux groupes sont issus du même environnement économique.

Afin de déterminer le groupe de contrôle comparatif à partir de l'enquête EDS, l'étude utilise une technique appelée le score de tendance. (Le fait que le questionnaire EDS soit exhaustif, pour ce qui concerne la collecte des données sur les caractéristiques des ménages, permettant d'arriver à un pronostic de la participation au programme, facilite l'utilisation de la technique du score de tendance). L'idéal, en matière de parité, serait deux individus, l'un appartenant à l'échantillon des participants, l'autre à celui du groupe de contrôle, pour lesquels toutes ces variantes (X) pronostiquant la participation au programme seraient identiques. Le problème courant, en matière de parité, est son manque de pragmatisme, au regard du grand nombre de variables incluses dans (X). Cependant, la parité peut être calculée à partir du score de tendance de chaque individu qui est tout simplement la probabilité de participation, fonction de (X). (Pour chaque observation au sein de l'échantillon des participants et du groupe de contrôle, le score de tendance est calculé en utilisant des modèles logit standards). Les données sur les revenus au sein du groupe de contrôle des non-participants permettent d'estimer les revenus renoncés par les participants à TRABAJAR II. Les revenus nets issus de la participation au programme sont ensuite calculés en tant que total des salaires du programme moins les revenus renoncés.

L'analyse cible est remarquable en ce sens qu'aucune collecte de données n'a été nécessaire. Le travail empirique s'inspire des données du bureau ministériel des projets à financer par département géographique pour TRABAJAR I (mars 1996 à avril 1997) et les six premiers mois de TRABAJAR II (mai à octobre 1997). Il s'inspire également d'un indice de pauvreté pour chacun des (510) départements, calculé à partir du recensement de 1991, en tant que proportion de ménages 'non satisfaits en termes de besoins essentiels'. Il s'agit d'un indice composite faisant état de la concentration urbaine, des installations sanitaires, de la qualité de logement, du niveau éducatif des adultes, de la fréquentation scolaire des enfants, de l'emploi et de la dépendance (rapport travailleurs / chômeurs, parmi les membres de la famille). L'indice est quelque peu obsolète, bien que cela ait l'avantage de rendre la mesure de la pauvreté départementale exogène (non-influencée), par rapport aux interventions TRABAJAR. Afin d'analyser l'incidence du ciblage, les données sur les dépenses publiques par zone géographique – dans le cas présent, les départements – sont régressées sur la base des taux de pauvreté géographique correspondants. Le coefficient qui en découle estime régulièrement un 'différentiel de ciblage' donné par la différence entre les allocations moyennes du programme en direction des pauvres et des moins pauvres. Ce différentiel national de ciblage peut ensuite être décomposé afin d'évaluer la contribution des mécanismes de ciblage du gouvernement central (les allocations de fonds au sein des départements) par rapport au ciblage au niveau provincial du gouvernement.

L'analyse des coûts et bénéfices a été effectuée par un ingénieur en génie civil qui a réalisé une étude en deux étapes des projets infra-structurels de TRABAJAR. Au cours de la première étape, elle a visité un échantillon de 50 projets TRABAJAR I terminés et les a notés sur la base d'indicateurs en six catégories : technique, institutionnelle, environnementale, socioéconomique, encadrement, et réalisation et maintenance. Les projets se sont ensuite vu attribuer une note globale sur la qualité, d'après un système de points, et des analyses des coûts-bénéfices ont été effectuées si nécessaire (sauf pour les écoles et les centres de santé). Une étude de suivi semblable pour 120 projets TRABAJAR II a été effectuée un an plus tard, visant à dépister l'impact des réformes sur la qualité de l'infrastructure.

Les évaluations sociales ont été réalisées au cours de la préparation du projet pour les programmes TRABAJAR I et TRABAJAR II. Ils fournissent des échos sur des questions de mise en œuvre des projets, telles que le rôle des ONG, la disponibilité de l'assistance technique dans la préparation et l'élaboration des projets, ainsi que la sélection des bénéficiaires. Les deux évaluations sociales ont été réalisées par des sociologues en ayant recours à des groupes d'intérêt et à des entrevues.

IV. Les Résultats

Prendre en compte les revenus renoncés est important pour pouvoir dresser un portrait fidèle des avantages du programme d'allocation conditionnelle décentralisée. Les statistiques descriptives des participants de TRABAJAR II suggèrent que, sans accès au programme, (revenus familiaux moins revenus du programme *par tête*), environ 85% des participants tomberaient dans la tranche des derniers 20%, relativement à la distribution des revenus nationaux et par conséquent, seraient classifiés comme pauvres en Argentine. Cependant, les estimations de la méthode de la comparaison appariée relative aux revenus renoncés sont considérables, si bien que les revenus nets moyens acquis à travers la participation au programme reviennent à environ la moitié du salaire sous TRABAJAR. Les participants au programme ne pouvaient pas se permettre d'être au chômage, en l'absence du programme. Il y a donc eu renoncement à une certaine partie des revenus, à travers la participation au programme. C'est cette partie des revenus renoncés qui est estimée en observant les revenus des non-participants 'comparés' à ceux des participants au programme. Cependant, même en prenant en compte les revenus renoncés, la distribution des gains est véritablement en faveur des pauvres, avec 80% des participants au programme tombant dans la tranche des derniers 20% dans la distribution des revenus. La participation au programme des femmes est faible (15%), mais les revenus nets sont pratiquement identiques pour les participants TRABAJAR, hommes et femmes ; les participants plus jeunes, cependant, obtiennent des revenus considérablement plus bas.

Le ciblage des performances s'est amélioré, de façon significative, consécutivement à la réforme de TRABAJAR II. Pour ce qui concerne l'allocation implicite des ressources en direction des ménages pauvres, celle-ci a été multipliée par sept, entre TRABAJAR I et TRABAJAR II. Un tiers de ce résultat est dû à un meilleur ciblage au niveau central, et deux tiers à un ciblage amélioré au niveau provincial. Il y a cependant des différences significatives dans le ciblage des résultats entre les provinces. Un département comportant 40% de sa population classifié en tant que pauvres peut s'attendre à recevoir jusqu'à 5 fois l'allocation départementale moyenne, selon la province à laquelle il appartient. De plus, ces performances de ciblage avaient tendance à être plus mauvaises dans les provinces les plus pauvres.

La qualité des projets d'infrastructure a été jugée adéquate, cependant, les réformes de TRABAJAR II, plutôt décevantes, n'ont pas abouti à des améliorations significatives. La raison en est partiellement la soudaine expansion du programme, qui a fait en sorte qu'il soit difficile, pour le programme, de répondre à certaines des normes de fonctionnement, lesquelles avaient été spécifiées *ex ante*. Néanmoins, les projets étaient mieux à même de répondre aux besoins prioritaires de la communauté. L'évaluation sociale fit ressortir le besoin d'améliorer l'assistance technique en direction des ONGs et municipalités rurales, ainsi qu'une plus grande publicité et une plus grande transparence de l'information à propos du programme TRABAJAR.

V. Application des politiques

Les résultats de l'évaluation apportent la preuve évidente que les participants au programme TRABAJAR sont très largement issus des couches pauvres de la population. L'auto sélection des participants par l'offre de bas salaires est une stratégie qui fonctionne en Argentine, et les participants obtiennent des revenus dans le cadre de leur participation (bien que ceux-ci soient plus bas que les salaires bruts, en raison des revenus renoncés). Les réformes de TRABAJAR II ont successivement amélioré les résultats de ciblage géographiques – le programme dès lors qu'il réussit mieux dans l'allocation des fonds en direction des zones pauvres ; cependant, les performances varient. Elles sont décidément faibles dans certaines provinces qui méritent plus d'attention en termes de politique. Finalement, des résultats décevants relativement à la qualité des projets d'infrastructure ont donné lieu à des efforts prodigieux de la part de l'équipe du projet, de façon à rehausser la performance dans ce domaine en améliorant les procédures de fonctionnement – en insistant sur la nécessité de procéder à un plus grand nombre de visites pour l'évaluation et l'encadrement, pénalisant les agences dont les performances, en matière de finalisation de projet, seraient en deçà des normes, et enfin en perfectionnant le manuel d'évaluation.

VI. Coûts d'évaluation et administration

Coûts. Les coûts relatifs à l'exécution de l'enquête TRABAJAR (relative à l'étude des revenus nets) et le traitement des données étaient d'environ 350.000 US\$. Les deux évaluations relatives à la qualité des avant-projets (analyses des coûts et bénéfiques) ont coûté à peu près 10.000 US\$ chacune, de même que l'évaluation sociale, ce qui amène le coût total des dépenses pour l'évaluation à environ 390.000 US\$.

Administration. L'évaluation a été conçue par un membre du personnel de la Banque Mondiale, un certain Martin Ravallion, et mis en œuvre, conjointement, par la Banque Mondiale et l'équipe argentine du projet. Au cours de ses différentes étapes, l'effort d'évaluation a nécessité une coordination avec plusieurs agences locales du gouvernement, y compris l'agence des statistiques, le Ministère du travail (y compris les bureaux de terrain) et la division de l'analyse des politiques du Ministère pour le développement social.

VII. Leçons tirées

Importance de la prise en compte des revenus renoncés dans l'évaluation des gains de l'allocation conditionnelle décentralisée. Les revenus renoncés représentent une portion considérable (environ la moitié) des revenus bruts gagnés par les participants au programme d'allocation conditionnelle décentralisée en Argentine. Les résultats suggèrent que les méthodes d'évaluation traditionnelles (qui font état, uniquement, du salaire brut) surestiment considérablement les revenus et par conséquent surestiment aussi la façon dont les participants pauvres s'en sortiraient sans le programme.

La méthode de comparaison par score de tendance. Lorsque la technique d'évaluation par comparaison appariée est utilisée, les scores de tendance permettent d'établir des parités fiables entre un échantillon de participants et de non participants (groupe de contrôle).

Utilisation judicieuse des sources de données nationales disponibles. Souvent, les sources de données disponibles telles que le recensement national ou les enquêtes sur les ménages peuvent procurer des informations précieuses relativement aux efforts d'évaluation. Se baser sur les sources disponibles réduit le besoin en collecte de données coûteuse dans le seul but de procéder à des évaluations. Les techniques d'évaluation innovantes peuvent compenser pour les données manquantes, comme l'indique très justement l'évaluation des résultats de ciblage géographique de TRABAJAR.

Une gamme étendue de composantes d'évaluation. La conception de l'évaluation TRABAJAR représente un mélange efficace d'outils et techniques d'évaluation. Les analyses des données de l'enquête, des visites de site et de l'évaluation sociale sont toutes utilisées afin de générer un grand nombre de résultats, lesquels représentent non seulement un apport précieux en termes d'efficacité du projet mais indiquent, de surcroît, les zones à réformer.

Pertinence des résultats. Plusieurs composantes de l'évaluation ont été conçues en gardant à l'esprit, de façon explicite, le cycle du projet, en des délais permettant d'obtenir des résultats au cours des étapes de préparation du projet, pour que les résultats puissent être utilisés efficacement lors de l'instruction de la politique. Plusieurs composantes donnent régulièrement lieu à des données dans un processus continu de suivi et de contrôle du projet.

VIII. Sources

Jalan, Jyostna, et Martin Ravallion. 1999. «Income Gains from Workfare and Their Distribution.» [*Revenus de l'allocation conditionnelle décentralisée et leur distribution*], Banque Mondiale, Washington D.C. Processed.

Ravallion, Martin 1999. Monitoring Targeting Performance When Decentralized Allocations to the Poor Are Unobserved. » [*Evaluation des performances de ciblage en l'absence d'observation d'allocations décentralisées en direction des pauvres*], Banque Mondiale, Washington D.C. Processed.

Annexe 1.2 : La Micro-finance aide-t-elle réellement les pauvres ? Preuves nouvelles à partir des programmes-vitrines au Bangladesh

I. Introduction

Description du projet. Les programmes de micro-finance de la Grameen Bank, Le Comité Pour la Promotion du Secteur Rural au Bangladesh, et l'Office du Développement Rural du Bangladesh sont des programmes-vitrines pour ceux qui ont été institués dans de nombreux autres pays. Ces programmes procurent des micro-crédits aux ménages pauvres, propriétaires de moins d'une acre de terrain¹. Les prêts sont accompagnés de contrats innovants et de plans de remboursement. Apparemment un succès, les programmes ont ainsi servi environ 4 millions de clients des milieux défavorisés du Bangladesh. Par exemple, le quartile supérieur des clients à l'emprunt de la Grameen Bank a une consommation supérieure de 15% ainsi qu'une proportion d'enfants mâles scolarisés deux fois plus importante. Il a aussi une proportion croissante de filles scolarisées, par rapport au quartile inférieur.

Points forts de l'évaluation. L'évaluation étudie l'impact du programme sur 1.800 ménages du Bangladesh et les compare avec un groupe de contrôle des ménages situés dans des zones dépourvues de financement de micro-crédits. La contribution principale de l'étude est de démontrer que des estimations simples de l'impact des programmes peuvent être considérablement exagérées : les revenus impressionnants sont apparemment rendus nuls par la correction des préjugés de sélection. L'évaluation démontre qu'une grande part des gains ayant été perçus diffère, selon les individus qui bénéficient d'un prêt : ils ont tendance à être plus riches et à travailler plus que les groupes de contrôle. Une fois que l'on utilise des techniques appropriées, l'emprunt n'a plus aucun impact sur la consommation, et les enfants en zones du programme sont en fait moins performants que les enfants des zones de contrôle. Le facteur-clé déterminant est que le prêt, dans le cadre du programme, n'a pas suivi les directives d'éligibilité – En fait, beaucoup d'emprunteurs disposaient de terres dont la superficie dépassaient le maximum autorisé une demi-acre.

L'évaluation utilise à la fois une technique d'enquête intéressante et des techniques économétriques appliquées de façon fort imaginative. Un autre aspect intéressant est que l'évaluation s'intéresse également à l'effet de l'impact sur la variance aussi bien que sur la moyenne des résultats, déduisant par là-même que le gain principal du programme est la réduction des risques plutôt que d'augmenter les résultats médians.

II. Thèmes de recherche et conception de l'évaluation

Les chercheurs s'intéressent à l'identification de l'impact des programmes de micro-finance sur la consommation individuelle enregistrée, par personne, la variance de la consommation enregistrée, le travail enregistré par adulte, lors des mois précédents, la variance du travail enregistré par adulte, les heures de travail des hommes adultes par mois, les heures de travail des femmes adultes, pour les mois précédents, le pourcentage de scolarisation des jeunes hommes (de 5 à 17 ans) et le pourcentage de scolarisation des jeunes femmes (de 5 à 17 ans).

L'évaluation est basée sur les enquêtes et couvre 87 villages sondés 3 fois de 1991 et 1992. Les villages ont été choisis de façon aléatoire à partir d'une liste de recensement et d'une liste de l'administration, de 5 sous-districts ayant servi de contrôle et 24 sous-districts dans lesquels les

¹ acre = moins de 40 ares

programmes ont été mis en œuvre. Les enquêtes ont été menées au sein de 24 ménages, dans chaque village.

Ceci a permis aux chercheurs de diviser les ménages en 5 types différents, selon les critères d'éligibilité relatifs à la possession d'une demi-acre de terrain. Il n'est pas inutile de reproduire le schéma, qui montre comment créer des variables auxiliaires caractéristiques de la typologie et de la façon d'appréhender les préjugés de sélection.

Village 1 : Avec programme		Village 2 : Contrôle	
A Non éligible [b=1 ;e=0 ;c=0]		Ménages disposant de plus d'1/2 acre	B Ne serait pas éligible [b=0 ;e=0 ;c=0]
C Éligible hors participation [b=1 ;e=1 ;c=0]	D Participants [b=1 ;e=1 ;c=1]	Ménages Disposant d'1/2 acre ou moins	E Serait éligible [b=0 ;e=1 ;c=0]

En comparant les résultats pour le groupe D avec ceux du groupe C, nous sommes confrontés à d'énormes problèmes de sélection : il est prouvé que les ménages du groupe C ne participent pas parce qu'ils ont peur de ne pas pouvoir rembourser. Si la possession de terres est exogène, les groupes C et D peuvent cependant être comparés au groupe E, du fait que la différence de résultat est tributaire du placement du programme et non de l'auto sélection. Ceci, bien sûr, n'est pas vrai, dans la mesure où il y a des différences dans les villages. Lorsque c'est le cas, (à cause, peut-être, de placement non aléatoire), alors il est préférable d'adopter une approche *différence à différence*. De cette façon, un évaluateur est en mesure de calculer les résultats médians pour C et D, ceux de A, et calculer ensuite la différence. De la même façon, la différence entre les résultats médians pour E et ceux pour B peuvent être calculés, et les différences *intra-villageoises* peuvent ensuite être comparées.

III. Données

Les chercheurs ont rassemblé des données sur 1.798 ménages ; 1.538 d'entre eux étaient éligibles pour participer, et 905 ont effectivement participé. Les résultats des enquêtes ont été collectés en 1991 et 1992 après les récoltes des trois principales saisons rizicoles. Les variables-clés intéressantes étaient la consommation par personne au cours de la semaine précédente, le montant des crédits perçus, la quantité de terres possédée, le travail fourni au cours du dernier mois ainsi que les caractéristiques démographiques. Une source de données secondaires sur la transaction des terres est également utilisée pour évaluer l'activité du marché foncier.

IV. Techniques économétriques

Parmi les techniques utilisées, trois semblent particulièrement intéressantes. La première est, en l'occurrence, l'utilisation de données administratives, pour vérifier les hypothèses-clés qui sont nécessaires à l'utilisation d'une stratégie conceptuelle de discontinuation de la régression : *l'exogénéité* de la propriété foncière. La seconde est l'utilisation fort judicieuse des techniques de représentation graphique non paramétriques pour décrire à la fois les probabilités en termes d'éligibilité et celles de se voir octroyer un prêt comme fonction de propriété foncière. Ceci est combiné à une très bonne discussion visant à identifier le moment approprié d'utiliser une stratégie conceptuelle de discontinuation de la régression – puisque l'analyse graphique suggère qu'il n'y a pas de point critique clair à 0.5 acre. Finalement, l'étude utilise principalement les techniques de différence et de différence-dans-les-différences.

V. Ceux qui l'ont exécuté

Les données ont été rassemblées par l'Institut des Etudes sur le Développement du Bangladesh, pour le compte de la Banque Mondiale. L'analyse a été effectuée par le chercheur Jonathan Morduch.

VI. Résultats

Les résultats suggèrent que presque tous les gains apparents qui sont issus du programme sont dus à des préjugés de sélection causés par une erreur de ciblage de prêt. En particulier, les auteurs constatent qu'entre 20 à 30% des emprunteurs possèdent plus de terre que le maximum d'1/2 acre requis dans le cadre du programme, ce qui indique que les responsables du programme sont susceptibles de contourner le règlement de façon indétectable. Lorsque les comparaisons sont limitées aux emprunteurs qui répondent aux critères de restriction foncière, les auteurs constatent également qu'avec les méthodes à la fois de la différence et de la différence des différences, les consommations moyennes au sein des villages ayant accès à la micro-finance sont inférieures à celles du contrôle. Cela indique qu'il y a eu des erreurs manifestes de ciblage des fonds du programme. Il en résulte que les approches de discontinuité de la régression ne peuvent pas être utilisées pour analyser les effets du programme.

L'évaluation est également utile dans le processus de comparaison des résultats à partir de techniques économétriques différentes : les résultats sont notoirement différents selon qu'on utilise des approches à effets fixes, de différence des différences ou, tout simplement de différences.

L'évaluation démontre de façon convaincante que le cas précédent est moins approprié lorsque des différences non-observées au sein des groupes cibles sont utilisées dans la prise de décision concernant la situation. Cependant on peut observer des résultats conflictuels dans les deux approches, selon que les programmes réduisent ou non la variation en termes de consommation et de revenus, faisant ressortir le besoin de données longitudinales. L'impact sur l'éducation, en fait, est inversé après correction sur les préjugés de sélection.

Il n'est pas inutile de souligner que malgré le fait que cette analyse fait ressortir un impact limité du traitement relatif au groupe de contrôle, celui-ci, en fait, peut ne pas avoir manqué d'accès au financement, car celui-ci pourrait bien avoir été procuré par les ONG. Les dépenses de millions de dollars pour subventionner les programmes de micro-finance, cependant, sont remises en question.

VII. Leçons tirées

Plusieurs leçons très importantes peuvent être tirées de cette étude, la première étant l'importance de vérifier si le programme fonctionne comme prévu. La deuxième est la question de la pertinence de la conception de la discontinuité de la régression, par rapport aux techniques de la différence des différences ou simplement de la différence. La troisième est relative à la considération de l'impact d'une intervention sur la deuxième ainsi que dès les premiers instants de la distribution, car la réduction des risques peut se révéler, à elle seule, être un résultat utile. Il y a une leçon plus fondamentale encore, qui, n'étant pas directement abordée, n'en est pas moins tirée de cette étude. Il s'agit d'une leçon d'économie politique : s'il y a de fortes incitations visant à faire une entorse à la loi, il en sera ainsi.

VIII. Sources

Morduch, Jonathan. 1998. « Does Micro-finance Really Help the Poor ? New evidence from Flafship Programs in Bangladesh. » Elaboré le 17 juin

Voir également :

Khandker, Shahidur R. 1998. Fighting Poverty with Microcredit : Experience in Bangladesh. New York : Oxford University Press for the World Bank.

Annexe 1.3 : Education contre vivres au Bangladesh : Evaluation d'un programme social ciblé lorsque le placement est décentralisé

I. Introduction

Description du projet. Le programme éducation contre vivres, au Bangladesh, (FFE) a été conçu comme une mesure incitative pour les parents, afin d'augmenter le taux de fréquentation à l'école primaire en distribuant du riz ou du blé à des ménages sélectionnés. Initialement prévu comme programme-pilote, il a pris de l'ampleur et de l'importance : sa part de budget de la Division de l'Éducation des masses et du primaire a augmenté de 11% en 1993-94 à 26% en 1995-96 pour atteindre 2,2 millions d'enfants, ou 13% du nombre total des fréquentations. La conception en est fort intéressante : le programme a été ciblé hiérarchiquement, en ce sens que FFE a été apporté à toutes les écoles des unités géographiques sélectionnées, en retard sur le plan économique et à faibles taux de scolarisation. Les ménages devant recevoir la nourriture ont ensuite été choisis par des groupes communautaires issus desdites unités géographiques, sur la base de critères établis, bien que relativement discrétionnaires, (ménages sans terre, ménages dirigés par une femme et ménages à bas revenus). Les enfants, dans ces ménages, doivent atteindre un taux mensuel de fréquentation des classes d'au moins 85%.

Points forts de l'évaluation. Cette évaluation est extrêmement utile car elle démontre ce qui peut être fait lorsque la conception de l'intervention n'est pas absolument favorable à des évaluations techniques standards et que l'évaluation doit être réalisée en utilisant des sources de données existantes. En fait, l'approche du FFE était totalement à l'opposé d'une mission complètement aléatoire : non seulement les zones géographiques ont été choisies parce qu'elles avaient certaines caractéristiques, mais les individus à l'intérieur de celles-ci ont été choisis parce qu'ils avaient besoin d'aide. Ainsi, puisque le programme ciblait les plus pauvres d'entre les pauvres, une analyse simple en minimisera forcément l'impact.

La conception de cette intervention constitue à la fois un problème important et un non-factuel car, à l'évidence, la sélection au programme est déterminée par le besoin des ménages. L'évaluation procure une approche innovante – et immédiatement transposable – pour s'attaquer aux préjugés, en se basant sur la décentralisation du processus de décision. En bref, du fait que le gouvernement alloue les fonds dans les zones géographiques, mais que les agents locaux procèdent à une allocation à l'intérieur de ces mêmes zones, l'évaluation utilise des techniques contributives variables basées sur la géographie afin de réduire le préjugé inhérent à la procédure de sélection endogène. L'application de la méthode résulte en une estimation beaucoup plus importante des impacts du FFE que des approches ordinaires.

II. Thèmes de recherche et méthode d'évaluation

Le thème de recherche est la quantification de l'impact du FFE sur la fréquentation scolaire, mesuré comme taux de fréquentation pour chaque ménage. Il n'y a pas grand chose en termes de méthode d'évaluation estimative : l'évaluation est réalisée grâce à des données déjà existantes – en particulier, en utilisant à la fois une enquête sur les dépenses des ménages représentative sur le plan national, et une enquête communautaire détaillée. L'évaluation rétrospective était en fait conçue pour rendre le besoin d'une enquête de référence évident ; l'évaluation nécessitait simplement des enquêtes prenant en compte les caractéristiques des ménages ainsi que les caractéristiques géographiques spécifiques de la zone du ménage. Les sections suivantes donnent plus de détail sur la façon dont ceci peut être structuré pour induire l'impact de l'intervention de façon fiable.

III. Données

Les données proviennent de l'Enquête (HES) sur les Dépenses des Ménages de 1995-96, une enquête représentative sur le plan national, menée par le Bureau Bangladeshi des Statistiques, qui comprend des questions sur la participation au FFE et dispose d'une composante d'enquête au niveau local. Les auteurs utilisent des réponses sur les caractéristiques démographiques des ménages, la propriété foncière, l'école et les variables de programme à partir de 3.625 ménages ruraux pour identifier l'impact sur la fréquentation scolaire. La fréquentation scolaire pour chaque enfant est en fait mesurée directement dans HES : les jours d'absence et les jours durant lesquels l'école est fermée sont comptés. La variable dépendante a été construite en tant que nombre moyen de jours de fréquentation scolaire par ménage, comme proportion du nombre de jours plausibles. Les deux parties de cette enquête sont décisives. D'un côté, les informations sur les ménages contribuent à saisir l'impact des caractéristiques démographiques sur la fréquentation scolaire. D'un autre côté, les informations sur les caractéristiques des emplacements géographiques permettent de modéliser la stratégie du processus décisionnel gouvernemental ainsi que de réduire les préjugés relatifs au processus de sélection mentionné ci-dessus.

IV. Techniques économétriques

L'évaluation s'occupe de deux problèmes très importants auxquels sont confrontés les chercheurs sur le terrain. En premier lieu, le placement du programme est décentralisé, par conséquent la décision d'allocation est tributaire de variables impossibles à observer par l'économétricien, et non par les décideurs. Cela signifie que l'évaluation nécessite une mesure qui détermine le placement du programme au niveau individuel sans être corrélé avec les termes de l'erreur (et donc les résultats du programme). En deuxième lieu il n'y a qu'une seule enquête transversale disponible, sans enquête de référence sur les participants de telle façon qu'il est plutôt difficile d'estimer l'impact absolu de l'intervention.

L'évaluation est extrêmement innovante en ce sens qu'elle utilise le processus d'allocation en deux étapes même comme un instrument. La caractéristique clé requise, en l'occurrence, serait que les données transversales comprennent à la fois des caractéristiques géographiques et des caractéristiques relatives aux ménages. Dans ce cas particulier, le modèle serait comme suit :

$$W_i = \alpha IP_i + \beta X_i + \gamma Z_i + \epsilon_i \quad (1)$$

Dans le cas présent, W représente le résultat individuel de l'allocation d'aide, X et Z comprennent les caractéristiques géographiques et relatives aux ménages, et IP , qui représente le placement individuel au sein du programme, est en corrélation avec les termes de l'erreur. Manifestement, et ceci est d'une importance fondamentale dans l'exposé relatif à l'évaluation, les estimations des moindres carrés de seront biaisées.

L'évaluation utilise les différences géographiques pour ce qui concerne le placement comme instrument pour le placement individuel, car il n'y a là aucune corrélation avec les termes de l'erreur, ainsi que les caractéristiques des ménages. Dans ce cas, cette relation est caractérisée comme suit :

$$IP_i = \delta GP_i + \pi X_i + \nu_i \quad (2)$$

Il est important de noter ici qu'il est crucial que Z contienne toutes les informations utilisées lors du processus décisionnel. Dans ce cas, les deux ensembles de variables géographiques sont utilisés. L'un des ensembles des variables géographiques est assez standard, affectant directement, à lui seul, les décisions de fréquentation : la distance de l'école, le type de l'école, ainsi que les variables qualitatives relatives à l'école. Le deuxième ensemble est relatif à la décision même de placement et, bien qu'il soit long, vaut la peine d'être mentionné ne serait-ce qu'à titre indicatif. Les variables comprennent la distribution foncière ; l'intensité de l'irrigation ; la qualité des routes ; l'électrification ; la distance et le temps pour se rendre dans le quartier général de l'administration locale, et à la capitale ; la distance vers les locaux des centres de santé et les institutions financières ;

l'incidence des désastres naturels ; l'attitude envers le travail des femmes, l'éducation et le planning familial; les niveaux moyens de scolarisation du chef de ménage et son épouse ; la religion majoritaire au sein du village, ainsi que la grandeur de la population du village. Ces calculs sont effectués au niveau du village et semblent assez bien prédire la sélection : Une régression probable (unités de probabilité) sur un total de 166 villages a donné des résultats relativement conformes (un pseudo- R^2 de 0.55). Ceci semble indiquer que ces variables saisissent bel et bien le placement global.

Cet ensemble d'équations peut être modelé en utilisant les moindres carrés en trois étapes. Il peut ensuite être comparé avec les résultats de la régression des moindres carrés ordinaires.

V. Ceux qui l'ont exécuté

L'évaluation a été réalisée par Martin Ravallion et Quentin Wodon, de la Banque Mondiale, dans le cadre d'un effort de collaboration à long terme, entre le Bureau Bangladeshi des Statistiques et l'Unité de Lutte Contre la Pauvreté et de Gestion Economique de la Région Asie du Sud de la Banque Mondiale.

VI. Résultats

Il y a des différences claires au sein des deux approches : l'impact estimé de FFE, en utilisant l'approche des moindres carrés en trois étapes était de 66% supérieur aux estimations des moindres carrés ordinaires sans contrôle géographique et 49% supérieur avec les contrôles. En d'autres termes, les estimations simples qui ne contrôlent que les variations au sein des ménages (les moindres carrés sans contrôle géographique), sous-estimeront l'efficacité du programme de façon substantielle. Même l'inclusion des contrôles géographiques pour apparemment contrôler le placement géographique n'efface en rien le préjugé présent. En termes substantifs la somme moyenne des gains au sein du programme semble avoir augmenté la fréquentation de 24% lorsque la méthode précisée ci-dessus a été utilisée.

Il est utile de préciser que le facteur-clé afin de faire de ceci une approche valide, réside dans le fait que suffisamment de variables sont disponibles pour modeler la décision de ciblage et que ces variables sont proches de celles utilisées par les administrateurs. S'il y a encore des variables omises, les résultats continuent d'être biaisés.

VII. Leçons tirées

De nombreuses évaluations n'ont pas le loisir de concevoir une stratégie de collecte de données à partir de la base, vers le haut, soit parce que l'évaluation n'était pas une partie intégrante du projet depuis le début, soit simplement pour des raisons de coût. Cette évaluation est importante à étudier pour deux raisons. D'abord, il informe sur le degré de préjugé susceptible d'avoir lieu dans le cas où une mauvaise approche économétrique serait utilisée. Ensuite, elle décrit une manière économétriquement viable d'estimer l'impact de l'intervention sans le coût et les délais impliqués dans une évaluation future.

VIII. Sources

Ravallion and Wodon. 1998. *Evaluating a targeted Social Program When Placement Is Decentralized*. Policy Research Working Paper 1945, World Bank, Washington D.C.

Annexe 1.4 : Evaluation du Fonds d'Investissement Social de Bolivie

I. Introduction

Description du projet. Le Fonds d'Investissement Social Bolivien (SIF) a été établi en 1991 comme institution financière dont la vocation est de promouvoir l'investissement durable dans les secteurs sociaux, plus particulièrement la santé, l'éducation et l'assainissement. Les objectifs de la politique sont de diriger les investissements dans des zones ayant été négligées depuis longtemps par les réseaux des services publics, plus particulièrement les communautés pauvres. Les fonds SIF, par conséquent, sont attribués selon un indice municipal de pauvreté, mais dans les municipalités, le programme est tributaire des demandes, en réponse aux demandes en projets des communautés, au niveau local. Les opérations SIF ont été décentralisées plus avant en 1994, amplifiant le rôle des ministères sectoriels et des gouvernements municipaux dans la conception et l'approbation des projets. Le SIF bolivien a été la première institution de ce type, dans le monde. Il a servi de prototype pour des fonds semblables qui, depuis lors, ont été introduits en Amérique latine, en Afrique et en Asie.

Evaluation de l'impact. Malgré la mise en œuvre généralisée des fonds sociaux, au cours des années 90, il n'y a eu que peu de tentatives, si rigoureuses soient-elles, d'évaluer leur impact sur la réduction de la pauvreté. L'évaluation du SIF bolivien, réalisée conjointement par la Banque Mondiale et SIF, a débuté en 1991 et se poursuit encore actuellement. L'étude comprend des données d'enquête de référence (1993) et de suivi (1997) lesquelles, combinées, permettent d'obtenir une évaluation d'impact 'avant et après. Elle comprend des évaluations séparées de l'éducation, de la santé et des projets hydrauliques. Elle est unique en ce sens qu'elle applique une palette d'évaluations techniques et examine les avantages et inconvénients de ces méthodologies alternatives. Les résultats initiaux de l'évaluation sont complets et sont actuellement présentés aux donateurs et agences gouvernementales pour rétroaction. Les résultats finaux ainsi que les questions méthodologiques seront examinés plus en détail relativement au rapport 2000 sur les Fonds Sociaux d'Investissement, avec une analyse de la rentabilité.

II. Conception de l'évaluation

Le processus d'évaluation du SIF bolivien a débuté en 1991 et se poursuit toujours. La conception comprend des évaluations séparées de l'éducation, de la santé et des projets hydrauliques, lesquelles évaluent l'efficacité du ciblage du programme en direction des pauvres, de même que l'impact des investissements de son service social sur les résultats communautaires souhaités, tels que les taux de fréquentation scolaire, les conditions sanitaires et une disponibilité de l'eau améliorés. Ceci illustre les techniques des meilleures pratiques dans l'évaluation, en utilisant des données référentielles dans l'analyse d'impact. L'évaluation est également innovante en ce sens qu'elle applique à l'analyse des projets d'éducation deux méthodologies d'évaluation alternatives – le procédé aléatoire³ et la comparaison appariée – et contraste ensuite les résultats obtenus selon chaque méthode. C'est une contribution importante car le procédé aléatoire est très largement considéré comme étant, statistiquement, la méthode la plus rigoureuse, cependant la comparaison appariée (utiliser un processus non-aléatoire dans la sélection d'un groupe de contrôle qui correspond le plus exactement aux caractéristiques des bénéficiaires du programme) est plus généralement utilisée dans la pratique.

² Sélection aléatoire des bénéficiaires du programme au sein d'un groupe éligible

III. Collecte des données et techniques d'analyse

Les efforts de collecte des données pour l'évaluation du SIF bolivien sont intensifs. Ils comprennent une enquête ('référentielle') sur l'investissement pré-SIF réalisée en 1993 ainsi qu'une enquête de suivi, réalisée, quant à elle, en 1997. Les enquêtes ont été appliquées à la fois aux institutions ayant reçu des fonds SIF et aux ménages et communautés qui bénéficient des investissements. Des données semblables ont également été collectées à partir d'institutions de comparaison (groupes de contrôle) et de ménages. L'enquête sur les ménages rassemble des données sur un ensemble de caractéristiques, comprenant la consommation, l'accès aux services de base ainsi que le niveau d'instruction et l'état de santé de chaque membre au sein d'un ménage donné. Il y a des échantillons séparés pour les projets de santé (4.155 ménages, 190 centres de santé), les projets d'éducation (1.894 ménages, 156 écoles), les projets hydrauliques (1.071 ménages, 18 projets hydrauliques) et les projets de latrines (231 ménages, 15 projets).

L'enquête sur les ménages consiste en trois sous-échantillons : (a) un échantillon aléatoire de tous les ménages en Bolivie rurale avec la région de Chaco (une province) ; (b) un échantillon de ménages vivant près des écoles, dans le traitement des groupes de contrôle pour les projets d'éducation ; et (c) un échantillon de ménages qui bénéficieront de projets hydrauliques ou de latrines publiques.

Afin d'analyser la mesure dans laquelle les investissements SIF sont efficacement ciblés en direction des pauvres, l'étude utilise les données référentielles (investissements pré-SIF) ainsi que l'information relative au placement ultérieur des investissements SIF, pour calculer la probabilité, pour certains individus, selon leur niveau de revenus, d'être bénéficiaires du SIF. L'étude combine ensuite les données de l'enquête référentielle et les données de suivi pour estimer l'impact médian de SIF sur les communautés ayant reçu un investissement SIF, en utilisant les techniques de régression. En plus de l'impact moyen, elle considère dans quelle mesure les caractéristiques des communes, des écoles ou des centres de santé peuvent être identifiées, une fois associées à des impacts considérablement plus importants que la moyenne.

Dans l'éducation, à laquelle les investissements SIF ont été aléatoirement attribués au sein d'un plus grand ensemble de communautés pareillement éligibles, l'étude applique la conception '*idéale*' de l'expérimentation aléatoire (dans laquelle le non-factuel peut être directement observé). Dans les projets de santé et d'assainissement, lorsque ceux-ci n'ont pas été attribués aléatoirement, l'étude utilise la méthode '*variable instrumentale*' afin de compenser le manque de non-factuel direct. Les variables instrumentales sont corrélées avec l'intervention, mais n'ont aucune corrélation directe avec les résultats.

IV. Résultats

Les investissements SIF II dans l'éducation et la santé donnent lieu à une amélioration manifeste en termes d'infrastructure et d'équipement. Les projets d'éducation ont peu d'impact sur les taux d'abandon scolaire, mais les scores de tests d'accomplissement scolaire parmi les 2^o secondaires sont bien meilleurs dans les écoles SIF. Dans la santé, les investissements SIF rehaussent les taux d'utilisation des services de santé tout en réduisant la mortalité. Les projets hydrauliques SIF sont associés à des améliorations très modestes dans la qualité de l'eau mais améliorent l'accès à l'eau potable ainsi que la quantité en réduisant également les taux de mortalité.

Une comparaison des résultats aléatoires par rapport aux appariés, dans l'éducation, démontre que l'approche appariée donne lieu à des groupes de traitement et de comparaison moins comparables et, par conséquent, à des résultats moins solides dans l'exercice de discernement de l'impact du programme. Pour illustrer cela, des preuves d'amélioration de l'infrastructure scolaire (que l'on s'attendrait à voir apparaître au sein des écoles SIF) sont décelées dans la conception de l'évaluation aléatoire, mais non dans la conception de la comparaison appariée.

Finalement, les résultats démontrent que les investissements SIF II ne sont généralement pas bien ciblés en direction des pauvres. Les projets santé et assainissement bénéficient à des ménages qui s'en sortent généralement mieux, en termes de revenus par personne. D'autre part, il n'y a aucune relation entre les revenus par personne et les bénéfices SIF dans l'éducation.

V. Application des politiques

Les résultats sur le ciblage font ressortir un conflit inhérent entre l'objectif de cibler les pauvres et le fait que le SIF soit basé sur la demande. Avec l'introduction de la loi sur la participation populaire de 1994, les sous-projets devaient être soumis par l'intermédiaire des gouverneurs municipaux. Les résultats du ciblage révèlent que même dans un système fortement décentralisé, il est capital de contrôler les processus de ciblage. Dans le cas bolivien, il semblerait que les communautés à la fois plus riches et mieux organisées, plutôt que les pauvres, sont celles qui sont plus susceptibles d'obtenir des investissements SIF. Dans le cas des projets d'assainissement SIF, en particulier, le préjugé contre les communautés les plus pauvres peut s'avérer être difficile à corriger. L'investissement dans l'assainissement de base est des plus efficaces dans les zones peuplées qui disposent déjà de l'accès à un système d'eau de façon à ce que le projet profite des économies d'échelle.

Le fait que les investissements SIF n'aient pas d'impact perceptible sur la fréquentation scolaire a déclenché une certaine restructuration des interventions SIF dans ce secteur. Plutôt que de se focaliser uniquement sur la fourniture d'infrastructures, les projets fourniront une combinaison d'extrants conçus pour améliorer la qualité scolaire. De même, des résultats décevants sur la qualité de l'eau (qui ne fait état d'aucune amélioration résultant des projets SIF par rapport aux sources préexistantes) ont suscité beaucoup d'attention. La conception des projets, dans ce secteur, est en passe d'être re-pensée.

VI. Leçons tirées

Efficacité de la technique d'échantillonnage aléatoire. La méthode de recherche aléatoire dans laquelle un groupe de contrôle est sélectionné de façon aléatoire à partir de bénéficiaires potentiels du programme est beaucoup plus efficace dans la détection des impacts du programme que la méthode de comparaison appariée visant à déterminer un groupe de contrôle. Le procédé aléatoire doit être inséré dans la conception du programme depuis le début, pour déterminer le processus à travers lequel les bénéficiaires du programme seront sélectionnés, et la sélection aléatoire n'est pas toujours faisable. Cependant, lorsque les fonds du programme sont insuffisants pour couvrir tous les bénéficiaires, on peut avancer un argument en faveur de la sélection au sein d'un plus grand nombre de bénéficiaires qualifiés.

L'importance d'institutionnaliser le processus d'évaluation. Les évaluations peuvent être longues et extrêmement complexes. L'évaluation bolivienne a été réalisée en sept ans, dans l'espoir d'appréhender les impacts du projet et de réaliser des résultats importants à cet égard. Cependant, l'évaluation a été difficile à faire pendant tout ce temps, particulièrement au regard de la gamme des différents acteurs impliqués (agences gouvernementales et institutions financières). La gestion et la mise en œuvre d'un effort d'évaluation peuvent être allégées en incorporant ces processus, dès le départ, dans le cours normal des activités ministérielles locales. Il serait préférable que d'autres efforts d'évaluation, plus intensifs, soient limités à un petit nombre de programmes – par exemple, des grands programmes dans lesquels il n'y aurait aucune réelle certitude concernant les résultats- et pour lesquels l'effort d'évaluation est susceptible d'être le plus payant.

VII. Coûts d'évaluation et administration

Coûts. Le coût total à ce jour de l'évaluation du SIF de Bolivie est estimé à US\$ 78.000, ce qui représente 0,5% du coût total du projet. La collecte de données représente la plus importante proportion du coût (69%), le reste des dépenses se répartissant sur les voyages, le temps d'intervention des employés de la Banque Mondiale et les consultants.

Administration. L'évaluation a été conçue par le personnel de la Banque Mondiale et financée conjointement par la Banque mondiale, l'Institut Bolivien des Statistiques Nationales et géré par les homologues SIF pour ce qui est de KfW et les gouvernements hollandais, suédois et danois. Le travail

d'enquête a été effectué pour ce qui concerne la première partie et ensuite, par le Ministère des finances pour la deuxième partie.

VIII. Source

Pradhan, Menno, Laura Rawlings, et Geert Ridder, 1998. «The Bolivian Social Investment Fund : An Analysis of Baseline Data for Impact Evaluation » *World Bank Economic review* 12 (3) : 457-82

Annexe 1.5 : Impact des programmes de travail actif: République Tchèque

I. Introduction

Description du projet. De nombreux pays en développement sont confrontés au problème du recyclage des ouvriers lorsque les entreprises contrôlées par l'état sont restructurées. Ceci est particulièrement compliqué dans les économies de transition qui sont également caractérisées par un taux élevé de chômage, et des salaires en stagnation ou en déclin. Cependant, tous les programmes de recyclage ne sont pas élaborés de la même manière. Certains ne sont que des paiements de subventions de licenciement déguisés pour des travailleurs déplacés; d'autres, des programmes de chômage camouflés. Ceci rend l'évaluation de tels programmes manifestement urgente.

Les programmes de formation, cependant, sont particulièrement difficiles à évaluer, et l'évaluation tchèque n'échappe pas à la règle. De façon caractéristique, plusieurs programmes différents sont institués pour servir différentes circonscriptions. Il y a également plusieurs façons de mesurer les résultats, y compris l'emploi, le travail indépendant, les revenus mensuels, et les revenus horaires. Plus qu'avec d'autres types d'évaluation, la magnitude de l'impact peut être sensiblement tributaire du temps : des résultats très différents peuvent être obtenus selon que l'évaluation a lieu un mois, six mois, un an ou cinq ans après l'intervention.

Points forts de l'évaluation. Cette évaluation a quantifié l'impact de quatre programmes sur le marché du travail actif (ALP) en République tchèque, en utilisant des méthodes de conception quasi-expérimentales – appariement des participants ALP avec un groupe similaire de non-participants. Les données d'enquêtes à la fois administratives et de suivi ont été utilisées au cours d'une évaluation ex post, pour divers résultats différents : durée de l'emploi, probabilité de l'emploi, emploi indépendant et revenus. L'analyse de régression est utilisée pour estimer l'impact de chacun des cinq programmes sur ces résultats, en surveillant les caractéristiques démographiques de référence.

Plusieurs leçons importantes ont été tirées de cette évaluation. Tout d'abord, un ensemble de leçons est de nature pratique : comment concevoir une évaluation particulièrement complexe, comment utiliser des données administratives, comment s'occuper des problèmes associés à l'administration de l'enquête et les mécanismes de création de l'échantillon apparié. Deuxièmement, comment structurer une analyse afin de fournir des informations sur les politiques – rendu possible par une évaluation détaillée de l'impact, par sous-groupe. Ceci a conduit à une recommandation sur les politiques pour cibler les programmes ALP vers des types de clients particuliers, et a conclu qu'un seul type d'ALP n'est pas du tout efficace pour changer soit l'emploi soit les gains.

II. Thèmes de recherche et méthode d'évaluation

Ceci se situe dans le cadre d'une évaluation plus large, pour quatre pays : La République tchèque, la Pologne, la Hongrie et la Turquie. Le contexte commun en est que chaque pays avait un taux de chômage important, dû, en partie, à la restructuration d'entreprises contrôlées par l'état, laquelle avait été réglementée par des programmes passifs d'aide aux revenus, tels que les allocations chômage et l'assistance sociale. Ceci avait été combiné aux ALP qui font l'objet de cette évaluation. Les cinq ALPs sont des Emplois Socialement Pertinents (nouvelle création d'emplois), des Emplois Publiquement Utiles, (emplois publics à court terme), des Programmes en direction des Jeunes sortant de l'Ecole (subventions pour l'embauche de jeunes diplômés), le Recyclage (formation spécifique pouvant durer quelques semaines à plusieurs mois), et les Programmes Pour les Handicapés. Cette dernière est plutôt réduite et n'est pas comprise dans l'évaluation. Il y a deux thèmes de recherche. L'un est d'examiner si les participants à différentes ALPs ont pu effectivement réintégrer le marché du travail dans une plus large mesure que les non-participants

et si cela varie au sein des sous-groupes, et selon les conditions du marché du travail. Le deuxième est de déterminer la rentabilité de chaque ALP et de faire des suggestions pour son amélioration.

L'évaluation est une méthode ex-post quasi-expérimentale – essentiellement une cohorte appariée. Le groupe de participants est apparié avec un groupe construit, de non-participants (en utilisant les informations tirées des archives administratives) sur des personnes enregistrées auprès des services de l'emploi de l'Etat, mais non sélectionnés pour l'ALP. La notion fondamentale est la suivante : un individu est sélectionné au hasard pour le groupe de participants ALP. Les résultats de cet individu sont ensuite comparés à ceux des individus au sein du groupe de non-participants (basé sur l'âge, le sexe, l'éducation, le nombre de mois sans emploi, la taille de la ville, l'état civil et la nature du dernier emploi). L'évaluation est particulièrement forte dans son analyse détaillée de la comparaison par rapport au groupe participant.

Il y a d'inévitables problèmes liés à cette approche. Ils ont été traités ailleurs, en profondeur (Burtless 1995, et Heckman & Smith 1995). Un problème évident, caractéristique de toute expérience non aléatoire est que les participants peuvent en fait avoir été 'écrémés' par le programme de formation sur la base de caractéristiques non observables ou impossibles à mesurer par les chercheurs. L'autre problème majeur est que les non participants peuvent avoir substitué d'autres types de formation à la formation publique, dans le cas du programme de recyclage. Le troisième problème est que les subventions pour embaucher les travailleurs peuvent tout simplement avoir donné lieu à la substitution d'un groupe donné de travailleurs à un autre.

III. Données

Une composante très intéressante de cette évaluation était l'utilisation de données administratives gouvernementales afin de créer le cadre de référence de cette enquête. L'équipe a ainsi rendu visite au ministère du travail et des affaires sociales (MOLSA) à Prague, et à trois agences locales de l'emploi pour comprendre à la fois le fonctionnement de l'administration et la mise en œuvre de l'ALP, ainsi que des données administratives des participants ALP. A partir de cela, 20 quartiers ont été choisis pour l'enquête, sur la base des critères de dispersion géographique et de variation des caractéristiques industrielles. Il y avait aussi une gamme étendue de taux de chômage, dans les différents quartiers. L'enquête contenait à la fois des questions quantitatives sur les résultats-clé du programme et des questions qualitatives, de la part des participants, sur leur estimation du programme.

Une autre composante précieuse était la mise en œuvre d'une enquête-pilote dans quatre quartiers. Cette approche, qui est toujours importante, a permis d'identifier, non seulement les problèmes techniques, mais aussi un problème juridique susceptible de survenir fréquemment, à cause de l'utilisation d'enregistrements administratifs. Il s'agit de l'interprétation de la loi sur la protection de la vie privée : dans ce cas précis, MOLSA n'a pas permis un mailing direct, mais demandait que les personnes concernées autorisent le bureau du travail à permettre la diffusion de leur adresse. Ceci a eu pour effet de retarder le calendrier de l'évaluation, d'augmenter les coûts et de baisser terriblement le taux de réponse.

L'enquête a été réalisée début 1997 sur un échantillon aléatoire de 24.973 individus ayant été contactés et enregistrés auprès de l'office du travail. Parmi ceux-ci, 9.477 ont participé à l'ALP au cours des années 1994-95. Le taux de réponse pour les non participants était de 14% ; pour les participants, il était de 24.7%, donnant un résultat global de 4.537 répondants. Le taux de réponse plutôt déprimant a été attribué directement à l'ordonnance juridique : la plupart des gens n'ont pas répondu à la demande initiale, mais parmi ceux qui ont autorisé la diffusion de leur adresse, le taux de réponse était important. Pire encore, le préjugé qui en résulte est inconnu.

IV. Techniques économétriques

La difficulté de mesurer à la fois la nature temporelle et la complexité des résultats du marché du travail est illustrée par l'utilisation de sept mesures de résultat différentes : le pourcentage actuellement employé, le pourcentage actuellement en emploi indépendant, le pourcentage jamais

employé, la durée de l'emploi, la durée de perception des indemnités de chômage, le total des paiements au titre du chômage et les revenus mensuels actuels.

L'approche d'évaluation, cependant, était tout à fait simple quant à son utilisation à la fois des différences simples entre les groupes et les moindres carrés ordinaires avec les pseudo groupes spécifiques pour jauger l'impact des interventions. L'impact global a été calculé, suivi des impacts estimés par chacune des catégories de sous-groupe (age, sexe, éducation, et, pour ce qui concerne les résultats des gains, la taille de l'entreprise). Cette dernière analyse était particulièrement utile car elle identifiait des sous-groupes d'individus pour lesquels, en fait, l'impact de l'intervention était différent, donnant lieu à des implications différentes sur les politiques. De fait, une recommandation majeure de l'évaluation était que les ALP soient ciblés de façon plus hermétique.

V. Ceux qui l'ont exécuté

L'évaluation a eu lieu dans le cadre d'un programme d'évaluation croisée des programmes de travail actif dans quatre pays, dans le but précis de comprendre l'impact des ALP sous différentes conditions économiques. L'évaluation a été supervisée par un comité d'organisation des projets avec des représentants de la Banque Mondiale, de chacun des quatre pays, des agences extérieures de financement ainsi que des consultants pour l'assistance technique (Abt Associates et l'Upjohn Institute).

Pour réaliser l'enquête, l'équipe a contracté avec une entreprise d'enquête privée, à la fois pour des raisons de qualité des données et pour réduire les possibilités d'intimidation dans le cas où l'agence de l'emploi locale devrait mener l'enquête elle-même. Il serait utile de noter, à ce propos, que la crédibilité de l'étude serait compromise dans le cas où le service de l'emploi serait responsable de la conduite de l'enquête. Effectivement, ce problème de danger moral est généralement important si l'agence responsable de la formation est également responsable de la collecte des informations sur les résultats de cette formation.

VI. Résultats

Les résultats sont caractéristiques des évaluations des programmes de formation. Quelques interventions semblent avoir certains impacts (si faibles soient-ils) pour certains types de travailleurs, dans certaines situations. Un point fort de l'évaluation est qu'elle identifie l'un des programmes qui semble avoir gaspillé de l'argent – aucun impact n'a été démontré, que ce soit globalement ou pour un sous-groupe. Un autre point fort est la présentation même de l'évaluation qui est particulièrement importante, si celle-ci doit être lue par les décideurs. Ici, les tableaux récapitulatifs des bénéfices combinés, en termes de salaires et d'emploi, globalement et pour les sous-groupes sont fournis pour chaque programme.

Un point très négatif, par contre, est que malgré la promesse initiale, aucune analyse des coûts n'a été réalisée. Il aurait été extrêmement utile de contraster les informations récapitulatives sur les bénéfices avec les coûts explicites et implicites combinés du programme. Ainsi, par exemple, bien que les évaluateurs aient remarqué que l'un des programmes augmentait la probabilité générale d'emploi, force est de remarquer que ceci n'a pu être réalisé qu'à l'issue d'un programme de neuf mois de formation. Un calcul total des taux de retour d'investissement aurait combiné le coût explicite du programme avec le coût d'opportunité du temps des participants, et l'aurait comparé avec l'augmentation en termes de revenus et d'emploi.

VII. Leçons tirées

Plusieurs leçons importantes ont été tirées de cette étude. Tout d'abord, il y a les composantes pragmatiques exposées dans l'introduction, particulièrement l'importance de prendre l'environnement politique en considération, dans la conception d'un plan d'évaluation. L'incapacité de convaincre le service de l'emploi de l'importance du projet d'évaluation a eu pour effet de compromettre sévèrement l'instrument d'enquête. Ensuite, l'étude démontre, de façon utile, la construction d'un échantillon apparié. Finalement, l'évaluation procure une bonne illustration de

l'importance de réaliser une analyse générale, et de sous-groupe, avec, comme résultante, la possibilité d'interventions ciblées fructueuses.

VIII. Sources

Benus, Jacob, Grover Neelima, Jiri Berkovsky, et Jan rehak. 1998. *Czech republic : Impact of Active Labor Market Programs*. Cambridge, Mass., et Bethesda, Md. : Abt Associates, May.

Burtless, Gary. 1995. «The Case for Randomized Field Trials in Economic and Policy Research» *Journal of Economic Perspectives* 9 (2) : 63-84.

Hechman, James J., et Jeffrey A. Smith. 1995. «Assessing the Case for Social Experiments.» *Journal of Economic Perspectives* 9 (2) : 85-110.

Schéma utilisé dans la conception du programme d'évaluation du travail actif tchèque



Annexe 1.6 : Impact du crédit avec éducation sur la nutrition des mères et de leurs jeunes enfants : Programme de banque rurale du 'Lower Pra' au Ghana

I. Introduction

Description du projet. Le programme Crédit et Education combine des éléments du programme de la Grameen Bank avec l'éducation, sur les notions de base en santé, nutrition, gestion et espacement des naissances et micro entreprise. L'objectif en est d'améliorer le statut nutritionnel et la sécurité alimentaire des ménages pauvres au Ghana. L'ONG '*Freedom from Hunger*', de pair avec le programme en nutrition internationale de l'Université de 'California Davis', a procuré du crédit et des services d'éducation aux femmes rurales démunies, dans le quartier est Shama Ahanta de la région Ouest du Ghana. Pour ces services, un partenariat a été établi avec cinq banques rurales – plus de 9000 prêts, totalisant 600.000 \$ ont été consentis jusqu'au mois de mars 1997 avec un taux de remboursement jamais inférieur à 92%.

Points forts de l'évaluation. L'évaluation est intéressante pour trois raisons : tout d'abord, la conception des échantillons était tout à fait appropriée : le programme était administré à 19 communautés et des données sur trois groupes échantillons de femmes différents ont été rassemblées : ceux ayant participé au moins une année, ceux n'ayant pas participé mais qui figuraient dans les communautés du programme, et ceux issus des communautés de contrôle. Deuxièmement, il y avait une description claire de son approche sous-jacente : elle a identifié et évalué à la fois les résultats intermédiaires et à plus long terme. Finalement, elle a fourni un mélange intéressant de résultats à la fois qualitatifs et quantitatifs et une illustration truffée d'anecdotes.

II. Thèmes de recherche et méthode d'évaluation

Le thème de recherche s'est focalisé sur les effets du programme sur (a) l'état nutritionnel des enfants ; (b) le pouvoir économique des femmes (revenus, épargne, temps) pour investir dans l'alimentation et les soins ; (c) la connaissance et l'adoption de l'allaitement, du sevrage et des pratiques courantes en prévention et gestion des diarrhées ; et (d) la capacité des femmes d'offrir à leurs enfants une alimentation saine et équilibrée.

Ce faisant, l'évaluation a séparé les objectifs finaux en termes de sécurité alimentaire des ménages et le statut nutritionnel des bénéficiaires intermédiaires du changement de comportement, de la réduction de la pauvreté et de la promotion féminine.

Une méthode quasi-expérimentale a été utilisée en effectuant deux enquêtes sur le terrain (en 1993 et 1996) pour évaluer l'impact de la stratégie sur le statut nutritionnel des enfants ; la capacité économique des mères, la promotion des femmes ainsi que l'adoption, par la mère, des pratiques en matière de santé infantile et de nutrition. Un total de 299 paires 'mères-enfants' a été sondé durant la première période et 290 paires différentes durant la deuxième période. Des informations à la fois qualitatives et quantitatives ont été rassemblées.

La conception de l'évaluation était plutôt complexe. La Banque rurale Pra a identifié 19 communautés n'ayant pas encore bénéficié des services crédit éducation, et les consultants ont divisé les communautés en grandes et en petites (plus petites ou plus grandes que 800 habitants) et encore une fois selon qu'elles étaient situées ou non près d'une route principale. Dans chaque stratification, les 13 ou les 19 communautés ont été assignées soit à un groupe de traitement, soit à un groupe de contrôle. Trois d'entre elles se sont vu attribuer un traitement pour des raisons politiques et trois autres communautés ont été sélectionnées en tant que contrôles appariés à celles ayant été sélectionnées pour des raisons politiques, sur la base de leur proximité, de leur développement commercial, de leur taille, et de leur accès à une route principale. Deux communautés se sont retirées par manque

d'intérêt à de petit nombre de communautés dans la classification. Ainsi, dans l'étude de suivi, seulement 17 communautés ont été sondées.

Dix paires mère-enfant, avec des enfants de 12 à 23 mois ont ainsi été choisies au sein de petites communautés, pour les enquêtes référentielles, et 30 dans des grandes communautés, en conséquence de quoi surgirent deux problèmes importants. Le premier est que cette construction n'a pas donné la possibilité à l'enquête de suivre, chronologiquement, la même femme, car peu de femmes, dans l'enquête référentielle, avaient aussi des enfants dans celle de 1996. Le deuxième problème est que la restriction, en termes d'âge, a tellement entaillé le deuxième échantillon, qu'il a été étendu à des femmes avec des enfants de moins de trois ans en 1996. Un avantage clé de cette méthode d'évaluation complexe était qu'il a été possible de classer les femmes, dans les échantillons référentiels, en futures participantes et futures non participantes.

Trois types de femmes ont été sondés : les participantes, les non participantes dans les communautés du programme, et les résidentes au sein des communautés de contrôle. Tous les participants ont été inclus ; les deux derniers types ont été sélectionnés au hasard parmi les femmes avec des enfants de moins de trois ans. Il est utile de mentionner que la taille totale de l'échantillon (360) a été calculée sur la base des déviations standards trouvées lors d'études précédentes, exigence que l'échantillon soit à même de détecter une différence de 0.4 de la valeur du score-z des groupes de contrôle et cibles, avec un niveau-cible de signification de 0.05 et une puissance de 0.8.

III. Données

Les données à la fois qualitatives et quantitatives ont été collectées, sur les ménages, les paires mère et enfant, se focalisant sur les mesures intermédiaires et à long terme – particulièrement la nature multidimensionnelle des résultats.

En ce qui concerne les résultats intermédiaires, cela amène à une série de questions visant à mesurer la capacité économique des femmes (revenus, profit, contribution aux revenus totaux du ménage, épargne, compétences en entreprise et dépenses en nourriture et ménage). De même, un autre ensemble de mesures se sont intéressés aux questions de savoir-faire des femmes en matière de santé et de nutrition (allaitement, nourriture infantile, prévention et traitement des diarrhées et immunisation). Pourtant, un autre ensemble de mesures a traité des questions de promotion féminine (assurance et espoir par rapport à l'avenir, statut et prise de décision au sein du ménage et le statut et les réseaux sociaux dans la communauté). Pour les résultats futurs tels que le statut nutritionnel et la sécurité alimentaire, des mesures plus directes ont été utilisées (mesures anthropométriques pour le statut nutritionnel et questions sur la faim pour la sécurité alimentaire).

Bien qu'un échantillon total de 360 paires 'mères et enfants' ait été planifié, seulement 299 paires ont été interviewées lors de la première enquête (principalement parce que deux communautés ont été abandonnées) et 290 lors de la deuxième. Les caractéristiques mères et ménages ont été comparées au sein des trois groupes ; aucune différence significative n'a été trouvée.

IV. Techniques économétriques

Les techniques économétriques utilisées sont plutôt simples. Elles ont exploité les points forts de la conception de l'enquête. La moyenne du groupe est calculée pour chacune des mesures variées ayant été utilisées, à la suite de quoi des tests-t sont faits pour voir si les différences entre contrôles et participants sont significatives. Il s'agit essentiellement d'une approche à différence simple. Celles-ci sont complétées par des graphiques.

Une série de questions clés, cependant, n'ont pas été traitées. D'abord, la conception de l'échantillon a été groupée – et parce que, presque par construction, les résultats de chaque paire individuelle mère-enfant seront corrélés avec les autres au sein de la communauté, les erreurs standards seront minimisées et les statistiques-t artificiellement maximisées. Dans les cas extrêmes, pour lesquels tous les résultats individuels sont parfaitement corrélés les uns avec les autres, la taille de l'échantillon est de 17 au lieu de 300. Ceci devra procurer une certaine signification à des résultats qui pourraient s'avérer en manquer. Deuxièmement, bien que la conception ait été explicitement stratifiée, l'impact de cette stratification n'a pas été traité : que les petites ou les grandes

communautés aient plus bénéficié ou que les communautés situées plus près d'une route soient plus riches que celles situées loin de toute route. Ceci est particulièrement surprenant, car l'on peut présumer que les raisons d'avoir une telle conception d'échantillon est bel et bien d'examiner les implications de la politique. Troisièmement, bien que les problèmes de préjugés de sélection soient discutés, il n'y a ni analyse formelle ni correction pour ce problème fondamental. Finalement, bien qu'il y ait eu des différences significatives au niveau des taux de non-réponse aux éléments, ce qui semble indiquer un préjugé potentiel au niveau de la sélection au sein même de l'enquête, ceci n'a jamais été traité ni même discuté.

V. Ceux qui l'ont exécuté

Un institut international à but non lucratif, '*Freedom from Hunger*' a développé le programme Crédit et Education et a collaboré à son évaluation avec le programme en Nutrition Internationale de l'Université de 'California Davis'. L'Institut est entré en partenariat avec la Banque Rurale Pra (une banque autonome, régulée par la Banque du Ghana) et ensuite avec 4 autres banques rurales ghanéennes, pour réaliser le programme. La Banque Rurale Lower Pra a joué un rôle dans le processus d'identification et de sélection des communautés à sonder.

VI. Résultats

Les objectifs intermédiaires ont généralement été atteints : bien que les revenus et les dépenses des femmes n'aient pas augmenté, les compétences d'entrepreneur des femmes ainsi que leur épargne étaient considérablement plus importantes. La connaissance des femmes en matière de santé et de nutrition s'est généralement améliorée. Les femmes étaient aussi plus susceptibles de se sentir responsabilisées. En termes de buts finaux, l'évaluation semble indiquer que le programme a effectivement amélioré l'état de la sécurité alimentaire des ménages et celui de la nutrition des enfants, mais pas l'état de la nutrition maternelle.

VII. Leçons tirées

Une contribution clé de cette évaluation est la conception très intéressante des échantillons : la stratification et le choix des groupes de participants et de non participants pour ce qui concerne leur participation future est une approche très utile. Une autre leçon est l'utilisation productive de nombreuses dimensions de résultats – quelquefois sur des facteurs plutôt non quantitatifs tels que la responsabilisation des femmes. L'autre leçon clé est la valeur des données quantitatives pour illustrer la validité des déductions quantitatives.

VIII. Sources

MkNelly, Barbara, et Christopher Dunford (en collaboration avec le Programme en Nutrition Internationale, Université de California Davis). 1998. « Impact of Credit With Education on Mothers' and their Young Children's Nutrition : Lower Pra Rural Bank Credit with Education program in Ghana. » *Freedom from Hunger Research Paper N° 4*, March.

Annexe 1.7 : Manuels scolaires et résultats des tests : Evidence à partir d'une éventuelle évaluation au Kenya

I. Introduction

Description du projet. L'évaluation des impacts de différents types de dépenses relatives à l'éducation sur les résultats des étudiants est particulièrement importante dans les pays en développement. Des études antérieures ont démontré que la fourniture de manuels scolaires est une manière efficace d'augmenter les scores lors de tests. Le Kenya, en raison des ressources extrêmement faibles dont disposent les éducateurs fait l'objet d'une étude fort intéressante dans ce domaine. Les évaluateurs remarquent que seul un enfant sur six, dans les classes de 3^o, de 4^o et de 5^o disposent d'un manuel scolaire. Dans les classes supérieures, le taux passe à un sur quatre. De plus, les infrastructures sont extrêmement déficientes, avec de nombreux enfants assis par terre pendant les cours.

L'évaluation a estimé l'impact sur les résultats scolaires en 1996, pour un programme dans lequel toutes les classes d'un sous-ensemble choisi au hasard, à partir de 100 écoles primaires rurales du Kenya, ont été dotées de manuels scolaires. Les manuels d'Anglais ont ainsi été distribués dans les classes des 3^o aux 7^o, avec un taux de 6 manuels pour 10 élèves, les manuels de mathématiques ont également été distribués aux classes de 3^o, 5^o et 7^o avec un taux de 50%, et les manuels de sciences aux classes de 8^o, avec un taux de 60%. En plus, chaque classe a été dotée d'un guide pour les professeurs. Des tests de performance ont été remis aux élèves avant à la distribution des manuels, puis 10 mois après. Les mêmes tests ont aussi été distribués aux écoles de contrôle. Cette approche combine une conception aléatoire avec des comparaisons réfléchies.

Points forts de l'évaluation. Cette évaluation est une illustration excellente de la façon de développer et de mettre en œuvre une bonne conception d'enquête puis d'effectuer un suivi avec des techniques économétriques appropriées. Elle est particulièrement efficace pour montrer comment faire des déductions sur les résultats de niveaux avec morcellement de données, l'utilisation d'estimations de différence dans les différences, comment s'attaquer aux préjugés de sélection et d'usage, de même que les erreurs de mesures et les questions de surpeuplement. Une autre composante très intéressante de l'évaluation est la focalisation sur les impacts de l'intervention sur les étudiants dans toutes les parties de la distribution. Finalement, la reconnaissance et l'analyse d'effets secondaires potentiels est un très bon exemple de comment appréhender toutes les dimensions d'une intervention.

II. Thèmes de recherche et méthode d'évaluation

Le point central de concentration de la recherche est l'évaluation de l'effet des manuels scolaires sur les résultats des études. Parce qu'il s'agit là d'un concept complexe, les résultats sont mesurés comme différence entre les écoles 'avec manuels' et les écoles 'de comparaison' et ceci dans plusieurs cas: les scores après tests, les gains de scores, les différences entre les combinaisons sujet-classe qui ont et qui n'ont pas reçu de manuels et les activités élève-enseignant. L'évaluation a également pris en compte d'autres effets secondaires (souvent ignorés), particulièrement la possibilité que la fourniture d'une telle subvention aurait pour résultat de réduire l'implication des parents, surtout dans l'abandon de la collecte de fonds.

La méthode d'évaluation est plutôt complexe. Le Ministère de l'éducation a choisi une centaine d'écoles nécessiteuses pour l'intervention de 1995. Celles-ci ont été réparties en quatre groupes – d'abord sur une base géographique, ensuite, alphabétiquement avec la géographie. Il y eut ensuite une tâche commandée, alphabétiquement, de chaque école, dans chacun des quatre groupes. L'assistance en manuels a été échelonnée pour atteindre le premier groupe, en 1996, le deuxième

groupe en 1997, etc. Les manuels de mathématiques, d'anglais et de sciences ont été fournis aux différentes classes – principalement classes de 3^o jusqu'en 7^o.

III. Données

Les examens de maths, d'anglais et de sciences ont été donnés aux enfants, dans toutes ces classes, à chacune des 100 écoles, avant la distribution des manuels. L'évaluation même, cependant, utilise des pré-tests ayant été administrés dans les classes de 3^o jusqu'en 7^o, en octobre 1996 et des post-tests en octobre 1997. Il y a donc des données sur quelque 8.800 étudiants (toutes classes confondues) pour chaque matière, dans les 100 écoles et un total de plus de 26.000 observations. Parce que 25 écoles ont reçu le manuel au cours de cette période, les étudiants de ces écoles deviennent le groupe 'manuel scolaire' ; les autres 75 forment le groupe de comparaison. En plus des scores des tests, des données ont aussi été rassemblées sur les finances des écoles et les méthodes pédagogiques.

Des informations sur l'utilisation des manuels en classe ont été collectées par des observateurs formés ayant visité chaque école et pris des notes succinctes sur huit activités de classe possibles, (allant des activités générales élève-professeur jusqu'à l'utilisation des manuels par les élèves et les professeurs). Ces notes couvrent 15 minutes et ont été utilisées pour arriver à des pourcentages de temps passé, sur chaque activité, par professeur et par élève, pour un total de 551 périodes scolaires, quatre à cinq étudiants, dans chaque classe, ont été interviewés par le personnel de terrain, lesquels ont rempli un questionnaire sur la base des réponses fournies.

Finalement, des données ont été rassemblées sur les finances des écoles sur la base d'un questionnaire de 1997 sur les écoles et les comités scolaires, portant sur les activités de recherche de fonds.

IV. Techniques économétriques

Il est important de noter les questions intéressantes générées par cette technique d'échantillonnage. Les scores des tests, au sein d'une école, sont susceptibles d'être corrélés les uns avec les autres, comme le sont les scores à l'intérieur des classes. De même, les scores des tests, pour différentes matières, traitées par le même enfant seront corrélés. L'intervention peut également être évaluée en termes d'impact sur les résultats du niveau d'apprentissage des étudiants ou sur leurs acquis académiques. En général, les effets d'une intervention devraient être cohérents avec les différentes techniques économétriques et les différentes manières de considérer les données ; c'était très certainement le cas ici.

L'évaluation commence par donner des estimations à partir d'une régression simple à niveau pseudo-variable, avec des pseudo traitements pour chaque combinaison classe-matière, avec des effets aléatoires par école, par classe et par matière (la variable dépendante est, en l'occurrence, le changement dans les scores, des pré tests aux post tests). Une caractéristique intéressante en est que les factices peuvent être combinés de différentes façons :

- Mettre en commun plusieurs classes afin d'estimer l'impact des manuels sur une matière
- Mettre en commun tous les scores pour estimer l'impact moyen des manuels sur une classe définie, et
- Mettre en commun tous les scores et toutes les matières pour estimer l'impact pondéré moyen des manuels pour toutes les classes et matières.

Manifestement, la structure des effets aléatoires varie avec chaque approche. L'évaluation, dans cette composante, est fort claire.

Ensuite, l'évaluation se poursuit avec une approche de différence dans les différences, qui est relativement élémentaire en ce sens qu'elle compare simplement les scores des pré et post tests entre les écoles de contrôle et celles de traitement.

La troisième approche, qui est sensiblement plus compliquée car elle exploite les variations intra-scolaires, mérite d'être discutée. La régression appliquée ici implique des scores de test régressés sur des factices qui considèrent si les étudiants étaient (a) dans une école à manuels et (b) dans une

combinaison matière-classe ayant reçu un manuel. Ceci réduit les problèmes introduits par une certaine hétérogénéité des écoles ainsi que des problèmes de sélection des échantillons – dans le dernier cas, parce qu'il saisit l'effet sur les scores, pour un même étudiant, selon que celui-ci ait reçu un manuel scolaire ou non. Il assume, cependant, que les scores, au sein des différentes combinaisons classe-matière peuvent être ajoutés et enlevés, et cette forte hypothèse peut être à l'origine des différents résultats de cette approche.

Un thème récurrent, dans les évaluations, est le désir d'appréhender non seulement l'effet médian de l'intervention mais aussi l'effet sur les sous-groupes de récipiendaires. Cette évaluation fournit une illustration très utile de l'utilisation des termes de l'interaction et de la régression quantile. L'approche précédente implique une interaction entre les scores de test initiaux et les manuels factices pour saisir les effets des manuels sur les meilleurs étudiants par rapport aux plus pauvres, en utilisant à la fois des valeurs vrais et instrumentales (les scores de test initiaux sont corrélés avec les termes de l'erreur, résultant d'un préjugé). La seconde approche qui implique l'utilisation de la régression quantile, est également utile et de plus en plus populaire. De façon plus spécifique, du fait que la régression des moindres carrés ne capture que l'impact médian du programme de manuels, les régressions quantiles permettent une variation des effets du traitement selon l'endroit où se trouve l'étudiant dans la distribution.

L'évaluation est aussi particulièrement forte, procurant l'application d'une méthode pour chercher les préjugés de sélection et d'usage. La source potentielle principale de problèmes dans cette intervention est la promotion des différentielles et des taux de répétition entre les manuels et les écoles de comparaison. Par exemple, les enfants peuvent être promus différenciellement de la classe 2 (une classe sans manuel) à la classe 3 (une classe avec manuel) dans les écoles à manuels. La promotion différentielle minimise les résultats des classes dans lesquelles les plus mauvais étudiants sont insérés et, éventuellement, maximise les résultats de celles dont ils proviennent. Ces deux effets ont été saisis dans l'évaluation en procédant à une ré-estimation du modèle de deux manières : se séparer des redoublants des deux ensembles d'écoles et se séparer des étudiants les plus mauvais de chaque classe. La fiabilité des résultats sous les deux approches a confirmé l'impact de l'intervention.

Finalement, en illustrant le fait de considérer l'importance des effets secondaires, l'évaluation a quantifié l'impact de la fourniture des manuels sur la recherche de fonds par les parents. Ils ont trouvé que l'intervention a monopolisé la contribution des parents – le montant de l'assistance non-ICS reçue, par école de comparaison, était de \$ 465 et par école-manuel scolaire, de \$267 (la valeur moyenne des manuels ICS était de \$ 485). Ils ont utilisé une analyse simple de régression, et examiné et confirmé l'hypothèse que les écoles plus petites bénéficiaient de plus de monopole que les plus grandes.

Ceux qui l'ont exécuté. Une organisation à but non lucratif hollandaise, International Christelijk Steunfonds, a financé le projet. L'évaluation a été réalisée par un professeur du Massachusetts Institute of Technology (Kremer) et deux économistes de la Banque Mondiale (Paul Glewwe et Sylvie Moulin). Certains coûts ont été supportés par la National Science Foundation et le World Bank Research Committee.

V. Résultats

Les résultats de cette évaluation contrastaient passablement avec les résultats d'autres évaluations d'interventions de manuels scolaires. Les résultats de base étaient qu'il n'y avait aucun impact significatif des manuels scolaires sur les résultats académiques médians, mais qu'il y avait, cependant, un effet significatif sur les meilleurs étudiants. Ceci était cohérent avec différentes techniques d'estimation et recouplement des données.

VI. Leçons tirées

La leçon la plus utile de cette évaluation, était l'importance d'utiliser différentes techniques économétriques pour vérifier la cohérence des résultats empiriques. Même si la collecte des données était fort proche de l'idéal, il est important que l'impact estimé de l'intervention reste plus ou moins le même avec différentes hypothèses économétriques et spécifications de modèle. L'application de

régression quantile et des termes de l'interaction était également une façon très utile d'analyser l'impact sur différents sous-groupes de la population. Finalement, il est capital de chercher et d'identifier les effets secondaires – dans ce cas précis, le potentiel de surpeuplement.

Source

Glewwe, Paul, Mickael Kremer, et Sylvie Moulin. 1998. «Textbooks and Test Scores : Evidence from a Prospective Evaluation in Kenya. » Development Research group (DECRG), Banque Mondiale, Washington, D.C. Traité.

Annexe 1.8 : Evaluation du projet de vulgarisation agricole au Kenya

I. Introduction

Description du projet. Le premier Projet de Vulgarisation Nationale (NEP-1) au Kenya a introduit le système de gestion Formation et Visite (T&V), pour les services de vulgarisation agricole, en 1983. Le projet avait le double objectif de développement institutionnel et de prestation de services de vulgarisation en direction des paysans, dans le but d'augmenter la productivité agricole. NEP-II, a suivi, en 1991, et visait à consolider les gains acquis sous NEP-I en augmentant le contact direct avec les paysans, en améliorant la pertinence des informations de vulgarisation et des technologies, en mettant à jour les compétences des employés et des paysans et en accroissant le développement institutionnel.

Evaluation de l'impact. La performance du système de vulgarisation kenyan a été controversée et fait partie du débat plus large sur la rentabilité de l'approche T&V de la vulgarisation. Malgré l'intensité du débat, le rôle important des services de vulgarisation agricole, au sein de la stratégie développementale de la Banque Mondiale pour l'Afrique, et le volume important des investissements réalisés, très peu de tentatives rigoureuses ont été faites pour mesurer l'impact de l'extension T&V. Dans le cas du Kenya, le débat a été élevé par des retours au T&V estimés très importants, rapportés dans une étude précédente, et le manque de résultats visibles convaincants, y compris la pauvre performance de l'agriculture kenyane, ces dernières années.

Le désaccord (entre le Département d'Evaluation des Opérations et la région africaine de la Banque Mondiale) à propos des performances de NEP-I persiste, en attendant cette évaluation, laquelle adopte une approche rigoureusement empirique pour évaluer l'impact du programme sur les performances agricoles. En utilisant le cadre de gestion basé sur les résultats, l'évaluation examine l'impact des services du projet sur la productivité et l'efficacité agricole. Elle développe aussi des analyses d'impact sur le programme (c'est-à-dire la conscience des agriculteurs et l'adoption de nouvelles techniques) et du rendement (par exemple la fréquence et la qualité des contacts) de façon à évaluer la performance du système de vulgarisation et de confirmer l'impact actuel ou potentiel.

II. Conception de l'évaluation

La stratégie d'évaluation illustre les techniques des meilleures pratiques en utilisant une palette de méthodes d'évaluation assez étendue, afin d'évaluer la mise en œuvre et le rendement du programme, ainsi que son impact sur la productivité et l'efficacité de la ferme. Aucune tentative n'est faite pour étudier l'impact sur l'assistance aux ménages, qui est susceptible d'être influencée par un certain nombre de facteurs n'étant pas du ressort des activités T&V. Elle fait appel à des méthodes à la fois quantitatives et qualitatives de façon à ce que des découvertes empiriques rigoureuses sur l'impact des programmes puissent être complétées avec des évaluations des bénéficiaires et des interviews de personnel devant mettre en relief des questions pratiques dans le processus de mise en œuvre. L'étude a également mis en application la méthode d'évaluation contingente pour clarifier les valeurs individuelles d'utilisation et de non-utilisation, pour une variété de publics et de marchandises et services publics. La méthode d'évaluation contingente fait ressortir les valeurs individuelles d'utilisation et de non-utilisation, pour une variété de publics et de marchandises et services publics. Les personnes interviewées sont appelées à déclarer leur disposition à payer (accepter) afin d'éviter un changement hypothétique dans la fourniture des biens ou des services – c'est à dire le résultat 'contingent'. Dans ce cas particulier, on a demandé aux agriculteurs combien ils seraient prêts à payer pour des services continus de vulgarisation agricole, dans le cas où le gouvernement cesserait de les procurer.

L'évaluation quantitative est compliquée par le fait que le système T&V a été introduit au niveau national, de façon à éviter une comparaison groupe de contrôle de types 'avec programme' et 'sans programme'. La méthodologie de l'évaluation a donc cherché à exploiter les données de la production agricole des ménages d'avant projet disponibles pour une comparaison limitée de type 'avant et après' utilisant des méthodes de données sur tableau. Pour cela, des données existantes sur les ménages ont été complétées par une enquête fraîche pour former un tableau. Les estimations des bénéficiaires conçus pour cette étude n'ont pas pu être réalisées, mais l'évaluation se base sur les découvertes pertinentes de deux estimations récentes effectuées au Kenya. L'étude est remarquable en ce sens qu'elle se base sur une palette de données pré existantes au Kenya (enquête sur les ménages, évaluations participatives, et ainsi de suite), complétées par un effort plus exhaustif de collecte des données allant dans le sens de l'évaluation.

III. Collecte des données et techniques d'analyse

L'approche de l'évaluation se base sur plusieurs sources actuelles de données quantitatives et qualitatives. L'évaluation quantitative est basée largement sur une enquête sur les ménages de 1998 réalisée par le Département d'Evaluation des Opérations de la Banque Mondiale. Cette enquête génère un ensemble de données en tableau en rendant une nouvelle visite à autant de ménages qu'il a été possible de localiser à partir d'une enquête sur les ménages de 1990 réalisée par le Département Technique pour l'Afrique, lequel, à son tour, s'est basé sur un sous-échantillon de l'Enquête sur le Budget des Ménages Ruraux de 1982. (Ces trois enquêtes génèrent un ensemble de données en tableau pour environ 300 ménages. L'enquête recouvre la démographie des ménages, les caractéristiques agricoles, et des données sur les intrants et les rendements de la production agricole ; les enquêtes de 1990 et de 1998 rassemblent également des informations sur les contacts avec les services de vulgarisation, y compris la conscientisation et l'adoption des messages relatifs à l'extension). Ces données sont complétées par une enquête du personnel de vulgarisation, plusieurs critiques récentes du service de vulgarisation, réalisées ou commissionnées par le ministère de l'agriculture, et des discussions individuelles ou de groupes d'intérêt avec le personnel de vulgarisation. L'étude se base également sur deux évaluations récentes des bénéficiaires : une étude, en 1997, par Action Aid Kenya qui a tiré au clair les opinions des utilisateurs et utilisateurs potentiels des services de vulgarisation du Kenya ; et une évaluation participative sur la pauvreté, en 1994, qui a enquêté sur les services publics, y compris l'extension, et a été réalisée conjointement par la Banque Mondiale, l'Administration du Développement Britannique à l'Etranger, La Fondation Africaine Médicale et de Recherche, l'UNICEF et le gouvernement kenyan.

L'analyse évalue à la fois le processus de mise en œuvre et les résultats du programme T&V au Kenya. L'étude évalue le développement institutionnel en se basant sur des données secondaires et qualitatives – enquêtes de personnel, interviews, ainsi que les examens des services de vulgarisation du ministère même. La qualité et la quantité des services fournis sont évaluées en utilisant une combinaison des découvertes des évaluations participatives (des bénéficiaires), des enquêtes de personnel, et à travers des mesures d'aide aux nécessiteux ainsi que la nature et la fréquence du contact entre les agents de vulgarisation et les agriculteurs, tirées des enquêtes OED¹ de 1998. Les données de l'enquête sont aussi utilisées pour mesurer les résultats du programme, en termes de conscientisation des agriculteurs et de recommandations sur l'adoption de l'extension.

Les résultats du programme – son impact réel sur la production agricole au Kenya – sont évalués en associant la fourniture des services de vulgarisation aux changements de productivité et d'efficacité au niveau de la ferme. Se basant sur les données en tableau sur les ménages, ces impacts sont estimés en utilisant l'analyse de l'enveloppement des données, une technique non-paramétrique, visant à mesurer les changements diachroniques dans l'efficacité et la productivité, de pair avec une analyse économétrique visant à mesurer l'impact de la prestation des services de vulgarisation sur la production fermière. Les méthodes d'évaluation contingente sont utilisées afin de mettre directement en évidence la volonté des agriculteurs de payer pour des services de vulgarisation.

¹ Département d'Evaluation des Opérations

IV. Résultats

Le développement institutionnel de NEP-I et NEP-II a été limité. Après 15 ans, l'efficacité des services de vulgarisation s'est très peu améliorée. Bien qu'il y ait eu, récemment, un remaniement au sein des approches de vulgarisation, le programme de vulgarisation a généralement manqué de vision stratégique pour un développement futur. La gestion du système est encore faible, et les systèmes d'information quasi inexistant. La qualité et la quantité de la prestation de services sont pauvres. Les bénéficiaires aussi bien que le personnel des services de vulgarisation nous rapportent que les visites sont à la fois inefficaces et trop peu fréquentes. Bien que les demandes en services techniques continuent à rester sans réponse, l'extension des services publics continue à se focaliser sur des messages agronomiques à la fois simples et basiques. Pourtant, l'approche adoptée – une forte intensité de contact avec un nombre limité d'agriculteurs – est adaptée pour dispenser de plus amples informations techniques. Le résultat a été un système de prestation de services coûteux et largement inefficaces. Les activités de vulgarisation n'ont eu que peu d'influence sur l'évolution des modèles de conscientisation et d'adoption des recommandations, ce qui indique un potentiel d'impact limité. Pour ce qui concerne l'impact réel sur l'efficacité et la production agricole, les données indiquent un léger impact positif des services de vulgarisation sur l'efficacité technique mais aucun effet sur l'efficacité économique. De plus, aucun impact significatif de la prestation de services sur la productivité, au niveau de la ferme, n'a pu être établi en utilisant simplement les données disponibles. Cependant, les données indiquent que l'impact a été relativement plus important dans les zones précédemment moins productives, où le fossé en termes de connaissance a probablement été le plus important. Ces découvertes sont cohérentes avec les découvertes de l'évaluation contingente. Une vaste majorité d'agriculteurs, à la fois parmi les récipiendaires et les non-récipiendaires actuels, sont disposés à payer pour recevoir des conseils, ce qui semble indiquer une demande restée sans réponse. Cependant, la valeur perçue du service, en termes du montant offert, est bien en deçà de ce que le gouvernement dépense actuellement pour le fournir.

V. Implications des politiques

L'évaluation des services de vulgarisation kenyans se singularise par la pléthore de conclusions relatives à la politique pratique qui découle de ces résultats, dont plusieurs sont cohérents avec les projets de vulgarisation agricole à venir. D'abord, l'évaluation révèle le besoin de rehausser le ciblage des services de vulgarisation, en se concentrant sur des zones et des groupes dans lesquels la différence entre la pratique générale et les meilleures pratiques est la plus importante. Par conséquent, l'impact est probablement le plus grand. De plus, les conseils doivent être minutieusement adaptés à la demande de la part des agriculteurs, en prenant en compte les variations des conditions économiques et technologiques locales. Afin de parvenir à ce niveau de ciblage des services avec succès, il est nécessaire d'obtenir un flux régulier et opportun d'informations fiables et appropriées, ainsi qu'un système d'évaluation et de contrôle pour assurer une rétroaction régulière des bénéficiaires sur le contenu des services.

Pour augmenter l'efficacité du programme, une présence plus discrète et moins intense des agents de vulgarisation mais avec une plus grande couverture sera probablement plus rentable. Il n'y a pas suffisamment d'innovations pour justifier des visites très fréquentes, et ceux qui sont actuellement privés d'accès exigent des services de vulgarisation. L'approche couverture du programme de prestations de services, laquelle repose principalement sur une seule méthodologie (visite de fermes) pour passer des messages standards simples, limite aussi l'efficacité du programme. Les programmes radio sont désormais populaires, les jeunes agriculteurs sont plus instruits et les prestataires alternatifs (organisations non-gouvernementales) commencent à émerger dans les régions rurales du Kenya. Une approche flexible, pluraliste de la prestation de services, et qui utilise plus particulièrement des moyens de communication peu coûteux, est certes susceptible de maximiser la rentabilité du programme.

Finalement, les principales découvertes indiquent le besoin d'une réforme institutionnelle. Comme avec d'autres services, une plus grande efficacité dans la prestation des services peut être obtenue grâce à des réaménagements institutionnels. Le point central de l'institution devrait être le

client (agriculteur). La décentralisation de la conception du programme, y compris des mécanismes participatifs donnant la parole aux agriculteurs (tels que le partage des coûts et les organisations de fermiers) devrait devenir une partie intégrale du mécanisme de prestation. La durabilité financière est capitale. La taille et l'intensité des services devraient être basées sur le fossé actuel, en termes de connaissance et de technologie, ainsi que l'afflux des nouvelles technologies. Le recouvrement des coûts, même partiel, offre plusieurs avantages : il procure des incitatifs appropriés, s'occupe de questions de responsabilité et de contrôle de la qualité, rend le service plus sensible et tributaire de la demande, et procure en quelque sorte une espèce de répit budgétaire. De tels réaménagements institutionnels décentralisés sont encore inexploités au Kenya, ainsi que dans plusieurs programmes de vulgarisation, en Afrique et dans le monde.

VI. Coûts d'évaluation et administration

Coûts. Le montant total du budget attribué pour l'évaluation était de \$ 250.000, qui ont couvert la collecte et le traitement des données de l'enquête sur les ménages (65.000 \$ – bien que ce soit probablement une sous-estimation des coûts réels) ; l'enquête du personnel de vulgarisation, et le rapport des consultants (12.500 \$) ; d'autres relatifs à la collecte des données (12.500 \$) ; et un analyste de recherche (8.000 \$). Approximativement 100.000 \$ (non rapportés dans les coûts officiels) des coûts en personnel pour la collecte, l'analyse et le rapport des données devraient aussi être ajoutés afin de refléter fidèlement le coût de l'étude.

Administration. Afin de maintenir une certaine objectivité et de dissocier le travail d'enquête à la fois du service de vulgarisation gouvernemental et de la Banque Mondiale, l'enquête sur les ménages a été mise en œuvre par l'Institut Tegemeo de l'Université d'Egerton, un institut de recherche indépendant au Kenya. L'analyse a été faite par Madhur Gautam, de la Banque Mondiale.

VII. Leçons tirées

- La combinaison de l'évaluation basée sur la théorie et un cadre de travail basé sur les résultats peut fournir une base solide pour évaluer l'impact des interventions du projet, particulièrement lorsque de nombreux facteurs sont susceptibles d'influencer les résultats attendus. La conception de cette évaluation a fourni des informations pour mesurer les indicateurs clés durant les étapes cruciales du cycle de projet, reliant les données du projet aux résultats attendus afin d'accumuler suffisamment de preuves d'impact.
- Une évaluation empirique exige une supervision intense et constante. Une évaluation peut être simplifiée de façon significative, avec un système d'estimation et de contrôle de grande qualité et fonctionnant correctement, spécialement avec de bonnes données référentielles. Des ressources adéquates pour de telles activités sont rarement disponibles. Cette évaluation a également énormément bénéficié du fait d'avoir accès à certaines données, si limitées soient-elles, au stade d'avant projet, ainsi que de sources indépendantes de données, à des fins comparatives.
- Une validation croisée des conclusions utilisant différentes approches analytiques et sources de données est importante pour rassembler un ensemble de preuves qui soit plausible. Des données imparfaites et des problèmes de mise en œuvre mettent certaines limites au degré de confiance que les méthodes individuelles peuvent inspirer, par rapport à des questions d'évaluation importantes. Des estimations quantitatives et qualitatives se complètent fort bien les unes les autres. L'expérience de cette évaluation indique que même en l'absence d'évaluation des bénéficiaires participants, des questions conçues de façon appropriée peuvent être incluses dans une enquête pour collecter à la fois des informations quantitatives et qualitatives. De telles informations peuvent procurer certains aperçus utiles à même de compléter des estimations quantitatives.
- L'évaluation contingente, si elle est correctement appliquée, peut être un outil utile, particulièrement pour ce qui concerne l'évaluation de la valeur d'un service public existant. Les résultats de

l'application, dans cette évaluation, sont encourageants, et les réponses semblent être à la fois rationnelles et raisonnables.

VIII. Sources

Banque Mondiale. 1999. *World Bank Agricultural Extension Projects in Kenya : An Impact Evaluation*. Operations Evaluation Department, report N° 19523. Washington, D.C.

Les titres suivants sont également disponibles au sein du Département Evaluation des Opérations de la Banque Mondiale :

*L'efficacité du système de vulgarisation agricole T&V au Kenya : Résultats d'une enquête sur les ménages
Conscientisation et adoption des messages de vulgarisation*

Reprise en considération des preuves sur les résultats de l'extension T&V au Kenya

*Efficacité des agriculteurs et changement de productivité au Kenya : Application de l'analyse
d'enveloppement des données*

*La volonté de payer pour les services de vulgarisation au Kenya : Application de la méthode d'évaluation
contingente*

Annexe 1.9 : Impact du Programme de Recyclage du Mexique sur l'emploi et les salaires (PROBECAT)

I. Introduction

Ce cas est quelque peu inhabituel en ce sens que le programme a fait l'objet de trois évaluations – d'abord par la Banque Mondiale qui a utilisé des données de 1992 (Revenga, Riboud et Tan, 1994) ; ensuite, par le Ministère Mexicain du Travail qui a utilisé des données de 1994 (STPS 1995) ; enfin, une mise à jour par la Banque Mondiale (Wodon et Minowa 1999). Les méthodologies utilisées pour les deux premières évaluations étaient assez identiques et ont donné des résultats identiques. Les consolidations méthodologiques de la troisième évaluation ont conduit à des conclusions et décisions de politique différentes. Le fait que les résultats diffèrent de manière substantielle entre les deux premières évaluations et la troisième souligne l'importance de la méthodologie et des données utilisées et la prudence dans l'interprétation des résultats lors de l'exécution des évaluations de programmes.

Description du projet. PROBECAT (Programa de Becas de Capacitacion para Trabajadores) est un programme mexicain de formation à court terme visant l'accroissement des revenus et l'emploi pour les travailleurs chômeurs et déplacés. Il est géré par le biais des bureaux publics d'emploi. Les stagiaires bénéficient d'un salaire minimum durant la période de la formation qui dure de un à six mois et les bureaux locaux de l'emploi sont responsables du stage. Au départ, le programme était de petite envergure (quelque 50 000 participants), mais au cours des dernières années il a connu une croissance rapide, comptant plus de 500 000 participants par an.

Points forts des évaluations. Les points forts sont les suivants:

- L'évaluation de 1994 est intéressante pour quatre raisons: l'utilisation géniale des données existantes; la mise en place d'un groupe de comparaison similaire; la reconnaissance explicite de la nature à contours multiples des résultats de l'intervention, en particulier pour les groupes hétérogènes de travailleurs; et l'analyse explicite de la rentabilité. Les conclusions de l'évaluation étaient assez positives en termes d'impact du programme sur les bénéficiaires.
- L'évaluation de 1995 est une reconduction de la méthodologie de l'évaluation de 1994 sur un ensemble de données plus récentes. Les conclusions sont également favorables pour l'impact du programme. Compte tenu du fait que la conception et les conclusions de l'évaluation de 1995 sont semblables à celles de l'évaluation de 1994, l'évaluation de 1995 ne sera pas discutée ici.
- L'évaluation de 1999 a eu lieu dans le cadre de l'évaluation de la pauvreté au Mexique en se basant sur l'ensemble des données utilisées dans l'évaluation de 1995 mais avec une méthodologie économétrique différente. Les contrôles utilisés pour l'endogénéité de la participation du programme ont montré la disparition de l'impact du programme sur la probabilité du travail et sur les salaires après la formation. Même si ce fait ne signifie pas que le programme n'a pas de profit, il suggère qu'il fonctionne plus en tant que réseau de sécurité temporaire pour les chômeurs que comme programme de formation pour l'emploi.

II. Thèmes de recherche et méthode d'évaluation

Dans l'évaluation de 1994, les auteurs mesurent l'impact de la formation sur (a) la probabilité de l'emploi après 3, 6 et 12 mois; (b) le temps de sortie du chômage; (c) l'effet sur les gains mensuels, les heures de travail hebdomadaire et les salaires horaires; et (d) le rendement sur l'investissement.

L'évaluation de 1999 étudie les mêmes questions à l'exception de celles sur les heures de travail hebdomadaire et les salaires horaires. Considérant qu'il n'existe pas d'impact dans cette

évaluation sur l'emploi et les salaires mensuels, le rendement est nul mais le programme peut fonctionner en tant qu'entité de sécurité.

La conception des deux évaluations est innovatrice par le fait de la mise en place du groupe de comparaison. Dans les deux cas, les évaluations impliquent l'entité d'enquête existante de la main-d'oeuvre, l'Encuesta Nacional de Empleo (ENEU) avec un groupe de stagiaires pour la même période. En d'autres termes, les critères de sélection du programme sont utilisés pour définir le groupe de contrôle de ENEU. Même s'il n'existe pas d'alternative à la combinaison des enquêtes pour cause de données limitées, la mise en place de l'échantillon conjoint (groupes de contrôle et de traitement) peut être mise en cause, comme il a été discuté dans l'évaluation de 1999:

- En utilisant des personnes sans emploi au niveau de ENEU pour former le groupe de contrôle, il est ressorti qu'aucune personne de ENEU n'a bénéficié du programme, ce qui n'est pas le cas car il existe pour chaque individu de ENEU la probabilité d'avoir participé à PROBECAT. Heureusement, vu que c'est un petit programme jusqu'en 1993, seule une petite minorité d'individus du groupe de contrôle ont bénéficié du programme (les données de l'évaluation de 1999 datent de 1993 – 94);
- La combinaison des deux échantillons pris au hasard (stagiaires de PROBECAT et individus chômeurs de ENEU) n'est pas un échantillon pris au hasard, de sorte qu'en l'absence de propriétés normales pour les restants, les résultats des régressions pourraient ne pas produire des estimations de paramètres consistants, en particulier parce que les modèles utilisés sont sensibles à la présomption de normalité bi-variée. En l'absence de meilleures données, peu d'action peut être entreprise dans ce domaine.

Les principales différences entre les évaluations de 1994 et de 1999 sont les suivantes:

- Dans l'évaluation de 1994, les auteurs tentent de se préoccuper des problèmes de parti pris dans la sélection résultant de la sélection au hasard des stagiaires par PROBECAT, en estimant un modèle de probabilité de participation. Le groupe de comparaison est donc limité aux individus qui sont plus susceptibles de participer. Dans l'évaluation de 1999, les auteurs soutiennent que cette méthode n'élimine pas le problème d'endogénéité de la participation au programme.
- Dans l'estimation des salaires dans l'évaluation de 1994, alors que la participation à PROBECAT est contrôlée, le parti pris dans la sélection de l'échantillon résultant de la décision de travailler n'est pas pris en compte. Dans l'étude de 1999, les deux problèmes de sélection de l'échantillon sont pris en compte.

III. Données

Dans l'évaluation de 1994, les données sur les stagiaires sont collectées à partir d'une enquête rétrospective de 1992 effectuée auprès de 881 hommes et 845 femmes formés en 1990. Elles sont complétées par des données d'une commission d'enquête auprès de 371 hommes et 189 femmes elles-mêmes obtenues d'une enquête sur les foyers de 16 zones urbaines principales du Mexique. Cette enquête faisait partie d'une enquête trimestrielle régulière sur la main-d'œuvre ENEU, entreprise par l'agence mexicaine des statistiques. Les auteurs ont profité de la structure de rotation du groupe de l'enquête pour prendre des travailleurs en chômage au cours du troisième trimestre de 1990 et ont ensuite suivi ces travailleurs pendant une année. Cela a été complété par un ensemble d'autres travailleurs ayant perdu leur emploi au cours du quatrième trimestre de la même année et suivis à leur tour pendant neuf mois. La même méthode a été utilisée dans l'évaluation de 1999, mais avec des données plus récentes.

IV. Techniques économétriques

Les principales techniques économétriques utilisées sont l'analyse de la survie (modèles de durée) pour la probabilité de trouver du travail et les régressions de Heckman pour les salaires.

Ce qui suit est basé sur l'évaluation de 1999. Les différences avec l'évaluation de 1994 sont mises en évidence.

Impact de PROBECAT sur la durée de recherche de l'emploi. Dans l'analyse de la survie, la fonction du survivant $S(t)$ représente la durée du chômage après la formation (mesurée en mois). Etant donné $S(t)$, la fonction de l'éventualité de rester au chômage $\lambda(t)$ dénotant la chance de trouver de l'emploi (ou le risque de rester au chômage) pendant le temps t pour les individus encore chômeurs en ce moment est $\lambda(t) = -d(\log S(t))/dt$. La courbe du survivant peut être spécifiée comme une fonction de participation du programme P , caractéristiques individuelles X et caractéristiques de l'état Z , de sorte que $\lambda = \lambda(t; X, Z, P)$. Dans le modèle de chances proportionnelles de Cox, si i dénote un foyer et j la zone où vit ce foyer, nous avons :

$$\lambda(t; X, Z, P1, P2) = \lambda_0(t) \exp(\gamma X_{ij} + \delta'Z_j + \mu P_{ij}). \quad (1)$$

Cox propose une estimation partielle maximale de l'éventualité de ce modèle dans laquelle la fonction de base $\lambda_0(t)$ n'a pas besoin d'être spécifiée. Si μ est positif et statistiquement important, le programme a un effet positif sur l'emploi. De manière stylisée, la différence entre les évaluations de 1994 et 1996 peut être décrite ainsi :

- Dans l'évaluation de 1994, les auteurs effectuent une enquête sur la probabilité de la participation du programme et retirent du groupe les individus avec une faible probabilité de participation au programme. Ils appliquent l'équation (1) sans contrôle supplémentaire quant à l'endogénéité.
- Dans l'évaluation de 1999, les auteurs ont également recours à des questions sur la participation au programme, mais ils profitent de l'existence du programme au niveau local (obtenue à partir des données administratives) comme facteur de participation supplémentaire déterminant (et non comme résultat conditionnel de la participation individuelle). Ensuite, ils appliquent l'équation (1), non avec la valeur réelle de la variable de participation mais avec la valeur (indice) attendue obtenue des questions initiales. Ceci est une procédure de variable instrumentale. L'idée est basée sur le travail effectué sur l'évaluation du programme utilisant les propriétés de décentralisation de Ravallion et Wodon (2000) et Cord et Wodon (1999). Les auteurs comparent leurs résultats avec ceux d'autres modèles, en montrant que d'autres méthodes font ressortir un parti pris dans la valeur des estimations du paramètre à cause du contrôle insuffisant de l'endogénéité.

Impact de PROBECAT sur les salaires mensuels. Pour effectuer cette analyse, un modèle avec contrôles pour la sélection de l'échantillon de la main-d'œuvre et la participation au programme est utilisé dans l'évaluation de 1999 (celle de 1994 contrôle seulement la participation au programme). Dénote par $\log w$ le logarithme du salaire attendu par individu. Ce salaire n'est pas nul si et seulement s'il est supérieur au salaire minimum accepté de l'individu. (Sinon, l'individu pourrait choisir de ne pas travailler). Dénote la différence non observée entre le salaire minimum accepté de l'individu et son salaire de réserve par Δ^* . Le salaire minimum accepté de l'individu est déterminé par un nombre de variables Z individuels (vecteur E , constitué essentiellement de l'éducation et de l'expérience antérieure de l'individu) et géographiques, plus la participation du programme P . La différence entre le salaire attendu de l'individu et son salaire minimum accepté est déterminée par les mêmes variables, plus le nombre d'enfants, le statut de chef de famille et la situation matrimoniale représentés par D . Le modèle est donc :

$$\Delta_{ij} = \phi \Delta^* E_{ij} + \Pi \Delta^* D_{ij} + \eta \Delta^* Z_j + \alpha \Delta P_{ij} + v_{ij} \text{ avec } \Delta_{ij} = 1 \text{ si } \Delta_{ij}^* > 0, \text{ et } 0 \text{ si } \Delta_{ij}^* < 0 \quad (2)$$

$$\log w_{ij}^* = \phi_w E_{ij} + \eta_w Z_j + \alpha_w P + k_{ij} \text{ avec } \log w^* \text{ si } \Delta = 1 \text{ et } 0 \text{ si } \Delta = 0 \quad (3)$$

Comme pour le modèle de la survie, pour vérifier l'endogénéité de la participation au programme, dans l'évaluation de 1999, des questions sur la participation au programme sont d'abord

posées en partant de la présence du programme sur le plan local comme facteur déterminant de la participation individuelle. Ensuite, les équations ci-dessus sont évaluées en utilisant la valeur (indice) projetée de la participation au programme au lieu de sa valeur réelle. Dans l'évaluation de 1994, le modèle ne s'assure pas de la décision de participer au marché de l'emploi donné dans l'équation (2) cidessus. Cette équation est remplacée par la probabilité de participation au programme comme variable indépendante. Là aussi, les comparaisons des différents modèles montrent l'existence de parti pris quand la variable technique instrumentale n'est pas utilisée.

V. Ceux qui l'ont exécuté

L'évaluation de 1994 a été effectuée par Ana Revenga du Département II Pays de l'Amérique Latine et des Caraïbes à la Banque Mondiale, Michelle Riboud du Département IV Pays d'Europe et d'Asie Centrale à la Banque Mondiale et Hong Tan du Département Développement du Secteur Privé à la Banque Mondiale. L'évaluation de 1999 a été effectuée par Quentin Wodon et Mari Minowa, également de la Banque Mondiale (Région Amérique Latine).

VI. Résultats

Les résultats obtenus au cours des diverses évaluations sont très différents. Les évaluations de 1994 et 1995 trouvent des impacts positifs du programme sur l'emploi et les salaires. Aucun impact positif n'a été trouvé dans l'évaluation de 1999 qui est basée sur les mêmes données utilisées dans l'évaluation de 1995. En termes d'analyse de la rentabilité, les deux premières évaluations sont favorables mais la dernière ne l'est pas. Les résultats défavorables de la dernière évaluation ne constituent pas une surprise. Il a été trouvé que la plupart des programmes de recyclage des pays de l'Organisation pour la Coopération Economique et le Développement ont un certain impact mais cet impact tend à disparaître au bout de quelques années (Dar et Gill 1998).

Le fait que PROBECAT ne soit pas probablement rentable pour les participants au long ou moyen terme selon la dernière évaluation ne signifie pas qu'il devra être supprimé. Le programme pourrait être perçu comme fournissant des réseaux de sécurité temporaires (par le biais de l'allocation du salaire minimum) plutôt que de formation. Il pourrait également être amélioré afin de fournir la formation avec des effets durables.

VII. Leçons tirées

Au-delà des quelques aspects innovateurs de ces évaluations et de leurs limites, la principale leçon apprise est la nécessité de prise de précaution dans les évaluations de programmes et de l'utilisation des résultats pour recommandation d'options de politiques. Le fait pour une évaluation plus récente utilisant des techniques économétriques différentes de pouvoir contredire une première devra être toujours gardé présent à l'esprit. Beaucoup de cas du genre existent dans les documents pour en attester.

VIII. Sources

Revenga, Ana, Michelle Riboud et Hong Tan. 1994. "The Impact of Mexico's Retraining Program on Employment and Wages." *World Bank Economic Review* 8 (2) : 247 – 77.

Wodon, Quentin et Mari Minowa. "Training for the Urban Unemployed : A Reevaluation of Mexico's PROBECAT." World Bank, Government Programs and Poverty in Mexico, Report No 19214-ME, Vol. II.

Annexe 1.10 : Programme national pour l'éducation, la santé et la nutrition (PROGRESA) au Mexique : Proposition d'Evaluation

I. Introduction

Description du projet. PROGRESA est un programme multisectoriel dont le but est la lutte contre l'extrême pauvreté au Mexique par le biais de la fourniture d'un ensemble intégré de services de santé, de nutrition et d'éducation, au profit des familles pauvres. Le gouvernement mexicain fournit l'assistance financière, les contributions nutritionnelles, les fournitures éducationnelles et une assistance de santé de base pour une période de trois années consécutives. Il compte élargir PROGRESA de sa taille actuelle de 400 000 familles de 1 à 1,5 millions de familles d'ici à la fin de l'année 1998 avec une dépense de 500 millions de dollars.

Points forts de l'évaluation. L'évaluation est particulièrement complexe parce que le programme est évalué sur trois dimensions: opération, recherche de l'efficacité et impact. Les résultats étant également multidimensionnels, ils ajoutent à la complexité. Ainsi, il existe plusieurs différents éléments d'évaluation: sélection des bénéficiaires, méthodes d'évaluation, cadre d'évaluation, d'analyse non expérimentale, besoins en données, impacts sur l'éducation, la santé, la consommation alimentaire et la nutrition, sur les dépenses en consommation et l'allocation intra-foyer, les impacts secondaires potentiels du programme, les simulations sur les changements dans les avantages du programme et les questions de rentabilité.

Bien que l'évaluation soit une ébauche d'idées plutôt que des résultats d'une exécution, la principale leçon apprise d'elle est la manière de penser et de structurer une évaluation avant son exécution réelle. En particulier, il existe une ébauche très utile des questions conceptuelles et empiriques à résoudre dans le cadre d'une évaluation et des voies par lesquelles les questions peuvent être étudiées. Un autre élément important de l'évaluation est son ampleur: au lieu de procéder à une simple évaluation de l'impact de l'intervention, il est utile de spécifier si le résultat a des chances de réussir dans l'opération et le ciblage de l'intervention.

II. Thèmes de recherche et méthode d'évaluation

L'essentiel des questions de recherche concerne l'évaluation des trois dimensions de la performance de PROGRESA – aspects opérationnels, ciblage et impact. L'aspect opérationnel d'une intervention est souvent ignoré en dépit du fait que les interventions peuvent être transformées d'échecs en réussites, si des mesures de correction sont prises. Un raisonnement similaire peut être fait sur le ciblage: un programme peut sembler avoir échoué simplement à cause d'un mauvais ciblage plutôt que parce que l'intervention a elle-même échoué. L'impact de l'évaluation est plus courant même si cet objectif est assez ambigu car tant l'ampleur de l'impact que les voies par lesquelles il est réalisé sont analysées.

Le contrôle de l'opération du programme est une procédure à deux niveaux. L'équipe met en place un schéma de la séquence des étapes de l'intervention. Ensuite, elle procède par observations, entretiens, groupes focaux et ateliers avec les acteurs pour évaluer, analyser et éventuellement changer les procédures du programme.

Une approche en deux étapes est également utilisée pour cibler les ménages devant participer au PROGRESA. La première est l'identification, au moyen d'un indice de pauvreté, des localités d'une région qui sont éligibles pour bénéficier de PROGRESA. La deuxième est l'identification de l'éligibilité d'une famille au niveau d'une localité, sur la base de l'interaction entre les responsables de PROGRESA et les élus locaux. L'étude se penchera sur la validité de ce ciblage en (a) comparant la distribution des niveaux de consommation des ménages parmi les ménages participants et non

participants dans les localités de traitement, (b) dérivant un découpage d'éligibilité pour la consommation des ménages conformément au nombre total de ménages pouvant être servis par PROGRESA, (c) effectuant une analyse de sensibilité et de spécificité des ménages PROGRESA et non PROGRESA contre les ménages sélectionnés dans le cadre de ce découpage, (d) exploitant la capacité des critères présents pour prévoir la consommation, (e) identifiant des critères de rechange à partir d'autres sources de données et (f) simulant des modèles susceptibles d'améliorer le ciblage par le biais de critères de rechange (Institut International de Recherche sur la Politique de l'Alimentation 1998, p.6).

Pour l'évaluation de l'impact, le même système a été suivi, avec le résultat que les localités ont été attribuées à 296 groupes de traitement et 173 groupes de non traitement, avec 14 382 familles dans la première catégorie et 9 202 familles dans la dernière catégorie. Les familles éligibles de la catégorie de contrôle bénéficieront de traitement après au moins une année.

Les consultants comptent tester la possibilité d'un choix au hasard en comparant les caractéristiques des groupes de traitement et de contrôle. Si ces derniers sont systématiquement différents, trois méthodes non-expérimentales seront utilisées: méthodes de contrôle du fonctionnement, d'assortiment et de régression.

III. Données

L'élément des données opérationnelles est obtenu de l'observation et des entretiens, groupes focaux et ateliers avec les acteurs. L'accent principal est mis sur la raison d'être et la manière dont les choses existent, le niveau de satisfaction quant au processus et les suggestions pour une amélioration. Ces données sont collectées dans toutes les localités et dépendront largement des archives administratives internes de PROGRESA.

Deux enquêtes ont été entreprises: les enquêtes de recensement de Décembre 1997 et les enquêtes de base de Mars 1998. La variable centrale pour le critère de ciblage est clairement la consommation des ménages; alors que cette dernière n'avait pas été prise en compte dans le recensement, elle l'a été dans l'enquête de Mars. Cette variable manque cependant de d'informations sur l'autoconsommation et même si elle va l'être plus tard, elle sera affectée par l'exécution de PROGRESA. Les consultants prévoient de travailler exclusivement avec les ménages éligibles et non éligibles au niveau des localités de contrôle.

L'évaluation de l'impact bute sur le choix des indicateurs d'impact. PROGRESA devra affecter à la fois la qualité et la quantité des services fournis et l'investissement en santé, nutrition et éducation. Un ensemble d'indicateurs d'évaluation est proposé sur la base de certains résultats d'impact et chacun provient d'une source de données associée. Le bien-être du ménage, tel que mesuré par la consommation du foyer, l'épargne, l'accumulation de biens durables sera mesurée par le biais d'enquêtes de base et de suivi; l'état nutritionnel et de santé des enfants sera mesuré par des enquêtes de sous-échantillons nutritionnels de base et de suivi; la réussite scolaire des enfants sera mesurée par des tests nationaux standardisés; la consommation alimentaire sera définie par des enquêtes de base et de suivi; l'utilisation de l'école sera étudiée à la fois par des enquêtes de base et de suivi et par une enquête au niveau scolaire; l'utilisation des structures sanitaires pourra être contrôlée à travers les archives des centres de santé et les enquêtes et le statut des femmes pourra être également mesuré par le biais d'enquêtes et les investigations des acteurs.

Un aspect très intéressant de l'évaluation proposée est l'approche analytique adoptée pour l'examen des mesures finales courantes et la discussion élargie des résultats et mesures des résultats plus appropriés pour l'éducation, la santé et la consommation.

Il est prévu une analyse de la rentabilité. Un ensemble d'avantages est mis en place, malgré la difficulté inhérente à la monétisation de la qualité de la vie et aux améliorations de la remise du pouvoir. Deux types différents de coûts sont également identifiés: coûts liés au programme administratif et coûts du programme. Les premiers concernent la sélection, les mécanismes de livraison ciblés et les coûts liés au contrôle. Les deuxièmes englobent la génération de revenus prévus.

IV. Techniques Économétriques

Les techniques économétriques appliquées dépendent des rapports à estimer. Les consultants discutent de la justesse du lien de la fonction de production (par exemple, pour la réussite académique), des liens du besoin (par exemple pour les services de santé ou d'éducation) et des liens de besoin conditionnel (pour lesquels certaines variables sont déterminées par la famille plutôt que par l'individu).

La technique économétrique la plus intéressante utilisée est appliquée à l'estimation d'une fonction de dépense de Working-Leser de la forme:

$$W_j = \alpha_i + \beta_{ij} I_{pcexp} + \beta_2 I_{siz} + \sum_k \delta_{kj} dem_k + \sum_s \theta_{sj} z_s + \beta_3 jP + e_j$$

où w_j représente la part du budget du nième bien; I_{pcexp} le log des dépenses totales per capita; I_{siz} , le log de la taille du ménage; dem_k la proportion du groupe démographique k dans le ménage; z_s un vecteur de variables fictives affectant la localisation du ménage; P représente la participation à PROGRESA et e_j les termes de l'erreur.

Cette approche offre beaucoup d'avantages: elle permet l'inclusion des facteurs de contrôle; elle répond à la contrainte de la contradiction et elle est largement utilisée, permettant des comparaisons avec d'autres études. Enfin, le modèle peut être utilisé pour identifier trois voies différentes à travers lesquelles PROGRESA est susceptible d'affecter les dépenses: à travers le changement des ressources des ménages (β_{ij} fois la tendance marginale à dépenser, estimée séparément), à travers le changement de la distribution du revenu (en le modifiant pour y inclure la proportion de femmes adultes dans le foyer) et à travers un effet de plus grande participation, les enquêtes de base et de suivi, l'utilisation des méthodologies des différences.

Elles identifient également les principaux problèmes économétriques susceptibles d'être rencontrés: co-linéarité, erreur de mesure, variables omises, simultanéité et identification de la période au cours de laquelle il est raisonnable de commencer à observer un impact.

V. Ceux qui l'exécuteront

Le personnel de l'Institut international de Recherche de Politique Alimentaire comprend Gaurav Datt, Laurence Haddad, John Hoddinott, Agnes Quisumbing et Marie Ruel. L'équipe est composée de Jere Behrman, Paul Gertler et Paul Schultz.

VI. Leçons tirées

La première leçon apprise ici est la valeur de l'identification des problèmes d'évaluation, la méthodologie et les sources – et l'évaluation critique de l'évaluation elle-même – avant le démarrage de celle-ci. Cette ébauche d'évaluation offre une opportunité très importante dans la mise en place d'une illustration réfléchie de tous les problèmes et manquements qu'un évaluateur est susceptible de rencontrer. En particulier, quelques questions de bon sens dans l'évaluation d'un impact sont identifiées: (a) les changements de politiques peuvent être difficilement prévus à cause de la substitution croisée et de l'ajustement de comportements; (b) les avantages et coûts marginaux dépendent d'un nombre d'éléments: externalité (fossé entre les évaluations sociale et privée), acteurs (parents contre enfants); (c) l'importance des caractéristiques non observées; (d) l'importance du contrôle des caractéristiques de l'individu, de la famille et de la communauté et (e) les estimations empiriques dépendant d'un environnement macro-économique, de marché, de politique et de régulation donné.

VII. Sources

Institut International de Recherche de Politique Alimentaire. 1998 Programa Nacional de Educacion, Salud, y Alimentación (PROGRESA); Une proposition d'Evaluation (avec annexe technique). Washington, D.C.: IFPRI.

Annexe 1.11 : Evaluation de la réforme scolaire au Nicaragua : Une approche quantitative et qualitative combinée

I. Introduction

Description du projet. En 1991, le Gouvernement du Nicaragua a introduit une réforme radicale de son système d'enseignement public. Le processus de la réforme a décentralisé l'administration scolaire (décisions relatives au personnel, au budget, au programme et à la pédagogie) et transféré les responsabilités financières à un niveau local.

Les réformes ont été mises en oeuvre sur une certaine période de temps, commençant en 1991 par un décret qui crée dans tous les établissements publics des conseils communauté-parents d'élèves. En 1993 un programme pilote au niveau de 20 établissements secondaires sélectionnés transforma alors ces conseils en comités de gestion avec des responsabilités plus importantes par rapport au personnel, au budget, au programme et à la pédagogie. En 1995, les comités de gestion scolaire étaient opérationnels dans 100 établissements secondaires et dans plus de 300 établissements primaires, inscrits au programme par un processus d'auto sélection impliquant un grand nombre d'enseignants et de directeurs d'établissements. Il était prévu que l'autonomie scolaire soit presque généralisée vers la fin de l'année 1999.

L'objectif des réformes du Nicaragua est de développer l'enseignement en changeant le processus organisationnel au sein des établissements publics afin que les décisions prises aient un impact positif sur les étudiants d'abord. Lorsque l'administration scolaire devient plus démocratique et participative et que les revenus générés localement augmentent, les procédures de dépenses deviennent plus rationnelles et consacrées aux efforts qui améliorent directement la pédagogie et augmentent la performance des étudiants.

Evaluation de l'impact. L'évaluation de la Réforme relative à l'Autonomie de l'Ecole au Nicaragua constitue l'un des premiers efforts visant à évaluer l'impact de la décentralisation sur les performances des étudiants. L'évaluation qui a été menée conjointement par la Banque Mondiale et le Ministère de l'Education a commencé en 1995 et devait être achevée vers la fin de 1999. La conception est innovatrice en ce sens qu'elle combine des méthodes d'évaluation qualitatives et quantitatives, et que la composante quantitative est unique en ce sens qu'elle intègre un module séparé d'évaluation des processus de décisions scolaires. L'évaluation illustre également les "meilleures" techniques à "utiliser" lorsqu'il n'y a pas de données de référence et lorsque l'application sélective (non randomisée) des réformes exclut une méthode d'évaluation expérimentale.

L'objectif de la composante qualitative de l'évaluation est de déterminer si la gestion et les réformes financières souhaitées sont réellement appliquées ou non dans les écoles et d'évaluer comment les attitudes des différents partenaires vis-à-vis de la réforme. La composante quantitative concrétise ces résultats en répondant à la question suivante: "Est-ce que les changements intervenus dans l'administration scolaire et le financement des écoles permettent aux enfants d'obtenir de meilleurs résultats scolaires?" Les résultats qualitatifs montrent que l'application réussie des réformes dépend en grande partie de la situation scolaire et de l'environnement (c'est à dire du degré de pauvreté de la communauté), alors que les résultats quantitatifs suggèrent que l'accroissement de la prise en charge par les écoles est en fait lié de façon significative à l'amélioration de la performance des élèves.

II. Méthode d'évaluation

La méthode d'évaluation relative à la réforme sur l'Autonomie de l'Ecole au Nicaragua est basée sur la "technique de comparaison appariée " dans laquelle les données pour un échantillon représentatif des établissements participants au processus de la réforme sont comparées aux données d'un échantillon d'établissements non participants. L'échantillon d'établissements non participants est sélectionné pour avoir des caractéristiques qui ressemblent, autant que possible, aux caractéristiques des établissements participants et fournir ainsi le non factuel. Cette méthode a été choisie parce que l'absence de données de référence a exclu toute possibilité d'emploi de technique d'évaluation à priori et à posteriori et parce que les réformes n'étaient pas appliquées aléatoirement aux établissements, ce qui a exclu une méthode expérimentale d'évaluation (dans laquelle l'échantillon des établissements qui ont fait l'objet de l'évaluation serait aléatoire et donc représentatif du point de vue national).

III. Collecte de données et techniques d'analyse

L'étude qualitative se rapporte à des données relatives à un échantillon de 12 établissements, dont 9 réformés et 3 non réformés, qui représentent le groupe de contrôle. Les données avaient en fait été collectées au niveau de 18 établissements, mais seuls 12 de ces établissements ont été inclus dans l'étude qualitative en raison des retards dans la préparation des relevés et d'une décision qui avait été prise de concentrer l'essentiel de l'analyse sur les établissements de la réforme, fournissant ainsi plus d'éléments significatifs à l'analyse. L'échantillon des 12 établissements avait été choisi être représentatif des établissements primaires et secondaires et des établissements urbains et ruraux, et était basé sur l'étude quantitative menée en 1995, et sur des établissements ayant des degrés différents d'autonomie dans la prise de décision. Au total 82 entretiens et des sessions de groupes spécialisés avaient été menés, mettant l'accent sur la question de savoir quelle signification les directeurs d'établissements, les membres du conseil, les parents et les enseignants donnaient au processus de décentralisation et quelle était leur attitude vis-à-vis de celui-ci. Tous les entretiens avaient été réalisés par des Nicaraguayens d'origine, entraînés à travers des simulations d'entretien et des tests pilotes à utiliser une série de questions types sans canalisation des réponses. Les entretiens avaient été enregistrés par moyen audio, transcrits, et enfin rendus sur deux à quatre pages de transcription, le tout analysé pour identifier les éléments particulièrement significatifs et les thèmes fondamentaux concernant tous les établissements et acteurs, ainsi que les questions qui ont surgi entre les établissements de la réforme et le groupe de contrôle.

La collecte de données quantitatives comprenait deux composantes, une série d'études d'établissements réalisée en deux étapes (Novembre-Décembre 1995 et Avril-Août 1997) et des tests de performance d'étudiants fréquentant ces écoles qui avaient été organisés en Novembre 1996. L'étude réalisée au niveau de l'école a permis de collecter des données sur les taux d'inscription, de redoublement et d'abandon, sur les ressources matérielles et humaines, la prise en charge au niveau des décisions, ainsi que les caractéristiques des directeurs d'établissements, des enseignants, des étudiants et de leurs familles. Le module de prise de décision au niveau des écoles est unique et présente une série de 25 questions conçues pour vérifier si et comment la réforme a réellement augmenté la prise de décision au niveau des établissements. L'étude a concerné 116 établissements secondaires (73 réformés et 43 non réformés représentant le groupe de contrôle) et 126 établissements primaires (80 réformés et 46 non réformés). Ici encore, les groupes de contrôle avaient été sélectionnés pour avoir les mêmes caractéristiques que les établissements réformés. L'enquête a également permis de collecter des données sur 400 enseignants, 182 membres de conseil, et 3000 étudiants et leurs parents, et 10 à 15 étudiants avaient été choisis au hasard au niveau de chaque école. Les étudiants qui sont restés à l'école et que l'on pouvait retrouver ont subi un test de performance à la fin de l'année scolaire 1996 et au cours de la deuxième série de collecte de données d'étude en 1997.

L'analyse des données quantitatives porte sur des techniques de régression pour estimer une fonction de production de l'enseignement. Cette technique permet d'examiner l'impact du système d'administration scolaire (son degré de décentralisation) sur les niveaux de performance des étudiants, le contrôle des performances des scolaires, et les caractéristiques des ménages et des

étudiants. L'analyse mesure à la fois les impacts de la décentralisation de droit et de fait; la décentralisation de droit indique simplement si l'établissement a été légalement ou non touché par la réforme, alors que la décentralisation de fait mesure le degré d'autonomie réalisé par l'établissement. La décentralisation de fait est mesurée comme étant le pourcentage de 25 décisions importantes prises par l'établissement lui-même et peut varier suivant les établissements parce que les réformes étaient organisées suivant des phases (de telle sorte que les établissements compris dans l'échantillon seront à des étapes différentes du processus de la réforme) et parce que la capacité de réussite de la mise en oeuvre des réformes varie selon le contexte de l'établissement (un résultat identifié dans l'étude quantitative).

IV. Résultats

L'étude qualitative montre que les changements de politique au niveau central n'ont pas toujours induit des changements conséquents au niveau local. En général, les réformes sont associées à une participation accrue des parents ainsi qu'à des améliorations au niveau de la gestion et du leadership. Mais le degré de réussite dans la mise en oeuvre des réformes est fonction du contexte scolaire. Sont considérés comme étant particulièrement importants le degré de pauvreté de la communauté environnante (au niveau des communautés pauvres, il est difficile d'accroître le financement local des établissements) et le degré de cohésion parmi les membres du personnel scolaire (lorsque des facteurs importants comme les enseignants eux-mêmes ne se sentent pas intégrés au processus de la réforme, la réussite de la décentralisation a été limitée). Les décideurs ignorent souvent les contextes locaux hautement variables dans lesquels les nouveaux programmes sont introduits. Les résultats qualitatifs montrent que dans la situation du Nicaragua, l'objectif d'accroissement des financements locaux pour les établissements est susceptible d'être en pratique voué à l'échec, surtout dans les milieux pauvres, et il est donc nécessaire de réfléchir de nouveau à cette situation.

L'étude quantitative renforce le résultat selon lequel les établissements réformés sont vraiment en train de prendre plus de décisions eux-mêmes, surtout en ce qui concerne les questions pédagogiques et les questions relatives au personnel. L'autonomie de droit—même en cas de signature de contrat de réforme par un établissement—n'implique pas nécessairement un degré plus important de prise de décisions au niveau des établissements scolaires, et n'a pas les mêmes impacts sur ces établissements. Le degré d'autonomie réalisé dépend du niveau de pauvreté de la communauté et de l'expérience de l'établissement dans le processus de la réforme. Les résultats de la régression indiquent que l'autonomie de droit a peu d'effets sur les performances des étudiants; mais l'autonomie de fait – le degré de décentralisation réelle réalisé par l'établissement – est associée de façon significative à une amélioration de la performance de l'étudiant. (Ce résultat est provisoire en attendant une analyse plus approfondie de la série de données qui est disponible maintenant.) De plus, les simulations indiquent que l'accroissement de la décentralisation a un impact plus important sur la performance des étudiants que les améliorations relatives à d'autres indicateurs de concentration type en matière de politique, tels que la formation des formateurs, la réduction de l'effectif des classes, et l'augmentation du nombre des manuels scolaires.

V. Application des politiques

Les résultats de l'évaluation ont démontré clairement que la Réforme relative à l'Autonomie des Etablissements Scolaires au Nicaragua a donné des résultats palpables. Les établissements réformés prennent en effet maintenant plus de décisions localement – la décentralisation est en train de se réaliser, non seulement en théorie – et le développement de la prise en charge au niveau local a pour résultat l'amélioration de la performance des étudiants.

Les résultats indiquent également les domaines pour lesquels les orientations peuvent être améliorées, et cela a eu pour résultat l'introduction par le Ministère de l'Éducation d'un certain nombre de changements dans le programme de réforme scolaire. Maintenant, le programme donne une plus grande importance au rôle des enseignants et à la promotion des aspects pédagogiques de la réforme. La formation des formateurs fait partie intégrante maintenant du programme, et l'élaboration d'un

conseil pédagogique est à l'étude. De plus, en réponse aux problèmes financiers des communautés pauvres, le Gouvernement a réalisé un programme de subvention sur la base d'un schéma de la pauvreté. Enfin, les impacts tangibles de cette évaluation ont amené le Ministère à incorporer une composante d'évaluation permanente dans le programme de la réforme.

VI. Coûts d'évaluation et administration

Coûts. Le coût total de l'évaluation était approximativement de 495,000 \$, représentant moins de 1.5 pour cent du crédit de la Banque Mondiale. (Ce total ne comprend pas le coût des équipes de la contrepartie locale au niveau du Ministère de l'Education). 39% du coût total de l'évaluation ont été alloués à l'appui technique fourni par des consultants externes, 35% à la collecte des données, 18% au travail du personnel de la Banque Mondiale, et 8% au déplacement.

Administration. L'évaluation a été réalisée conjointement par le Ministère de l'Education du Nicaragua et la Banque Mondiale. Au Nicaragua l'équipe d'évaluation a été dirigée par Patricia Callejas, Nora Gordon, et Nora Mayorga de Caldera au Ministère de l'Education Nationale. A la Banque Mondiale l'évaluation a été réalisée dans le cadre du projet de recherche, "Evaluation d'Impact des Projets Education Concernés par la Décentralisation et la Privatisation" sous la direction de Elizabeth King avec Laura Rawlings et Berk Ozler. En coordination avec l'équipe de la Banque Mondiale, Bruce Fuller et Magdalena Rivarola de la Faculté des Sciences de l'Education de Harvard ont travaillé avec Liliam Lopez du Ministère de l'Education du Nicaragua pour réaliser l'évaluation qualitative.

VII. Leçons apprises

Valeur de l'approche de méthodes mixtes. L'utilisation à la fois de techniques de recherche qualitative et quantitative a produit une combinaison précieuse de résultats utiles et appropriés. Le travail quantitatif a fourni un aperçu riche et statistiquement valable des conditions et réalisations scolaires; le travail qualitatif a mis en valeur ces résultats avec des éclaircissements sur les raisons de la concrétisation de certaines attentes du programme de la réforme et de la non réalisation d'autres attentes, ce qui a contribué à guider les réaménagements nécessaires en matière d'orientation. De plus, en raison du fait que le travail qualitatif est plus intuitif, il était plus accessible et donc plus intéressant pour le personnel du Ministère, ce qui a, à son tour, facilité le développement rapide des capacités et de la crédibilité pour le processus d'évaluation au sein du Ministère.

Importance du développement des capacités locales. Le développement des capacités locales a été coûteux et a nécessité une fréquence au niveau des contacts et de la coordination avec des homologues de la Banque Mondiale et des consultants externes. Cependant, l'avantage a été le développement rapide de l'appropriation et de la responsabilité vis-à-vis du processus d'évaluation, ce qui a ainsi engendré un degré remarquable d'acceptation des résultats de l'évaluation, qu'ils aient influencé positivement ou négativement le programme. Ces résultats d'évaluation ont fourni des intrants directs à la réforme alors en cours. L'impact politique de l'évaluation a également été mise en valeur grâce à une équipe locale cohésive dans laquelle les évaluateurs et les décideurs ont travaillé en parfaite collaboration, et parce que le Ministère de l'Education a été impliqué au premier plan comme partie intégrante d'appui au processus d'évaluation.

VIII. Sources

Les documents suivants donnent des informations détaillées sur l'Evaluation de la Réforme Relative à l'Autonomie des Etablissements Scolaires au Nicaragua: Fuller, Bruce, and Magdalena Rivarola. 1998. *Nicaragua's Experiment to Decentralize Schools: Views of Parents, Teachers and Directors*. Working Paper Series on Impact Evaluation of Education Reforms, paper no.5. World Bank, Washington, D.C.

King, Elizabeth, and Berk Ozler. 1998. *What's Decentralization Got to Do with Learning? The Case of Nicaragua's School Autonomy Reform*. Working Paper Series on Impact Evaluation of Education Reforms, paper no.9. World Bank, Washington, D.C.

King, Elizabeth, Berk Ozler, and Laura Rawlings. 1999. *Nicaragua's School Autonomy Reform: Fact or Fiction?* Washington, D.C.: World Bank.

Nicaragua Reform Evaluation Team. 1996. *Nicaragua's School Autonomy Reform: A First Look*. Working Paper Series on Impact Evaluation of Education Reforms, paper no.1. World Bank, Washington, D.C.

Nicaragua Reform Evaluation Team. 1996. *1995 and 1997 Questionnaires, Nicaragua's School Autonomy Reform*. Working Paper Series on Impact Evaluation of Education Reforms, paper no.7. World Bank, Washington, D.C.

Rawlings, Laura. 2000. "Assessing Educational Management and Quality in Nicaragua." In Bamberger, *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C.: World Bank.

Annexe 1.12 : Améliorer l'enseignement des Mathématiques élémentaires au Nicaragua : Une étude expérimentale de l'impact des manuels et de la radio sur la performance

I. **Résumé de l'évaluation**

La plupart des pays pauvres ont des ressources extrêmement limitées pour faire face à l'enseignement, ce qui rend importante la répartition efficace de ces ressources. Des trois options disponibles les plus connues en matière d'orientation – effectifs non pléthoriques des classes, programmes de formation des formateurs plus longs, et provision de manuels – seule la dernière option a été fréquemment considérée comme ayant un effet réellement positif sur l'apprentissage des étudiants. Cette évaluation a quantifié l'impact de la disponibilité de manuels sur l'apprentissage des mathématiques chez les étudiants nicaraguayens de première année.

La méthode d'évaluation prévoyait la distribution de manuels à tous les étudiants dans un sous-groupe de classes qui avaient été à l'origine sélectionnées pour constituer le groupe de contrôle dans le cadre d'une étude en cours sur l'efficacité de programmes éducatifs à la radio. Une moitié des classes avait reçu des manuels; l'autre n'en avait pas reçu. Un pré-test avait été administré à toutes les classes au début de l'année et un post-test avait été administré à la fin de l'année. Des techniques simples de régression avaient alors été utilisées pour comparer la moyenne des résultats obtenus par les classes pendant le post-test en tant que fonction de résultat du pré-test et de l'intervention.

La principale leçon tirée est comment bien concevoir une évaluation: la sélection aléatoire a été bien faite et savamment combinée avec un test qui a mis en valeur la comparaison entre les différentes classes. Une autre leçon tirée était une leçon pratique: l'évaluation avait été conçue pour prévenir éventuellement des questions relevant de l'économie politique. Finalement, l'évaluation a fourni une série d'exemples pratiques relatifs aux types de décisions qui peuvent être prises sur le terrain.

II. **Thèmes de recherche et méthode d'évaluation**

Il y a deux composantes très intéressantes pour la méthode d'évaluation: le soutien énorme d'une évaluation précédente et la compréhension d'emblée du contexte politique dans lequel l'évaluation devait se dérouler. La préoccupation principale était tout d'abord: d'évaluer l'impact d'une meilleure disponibilité de manuels sur l'apprentissage des étudiants de la première année — en mettant l'accent particulièrement sur la question de savoir si les manuels étaient réellement utilisés en classe. En raison du fait qu'il y avait déjà un programme éducatif à la radio (Radio Mathématiques) en cours, la question avait été élargie pour estimer l'impact de la disponibilité des manuels par rapport à l'enseignement par la radio et au groupe de contrôle.

Il est utile de discuter de la décision de surveiller l'utilisation effective des manuels, ce qui rend l'évaluation plus difficile. Dans de nombreux cas d'interventions dans le domaine de l'éducation, du matériel est fourni aux classes, mais il est clair que l'impact de la fourniture de ce matériel dépend de l'utilisation. Cependant, comme l'ont indiqué les évaluateurs, cette décision signifie que l'évaluation "ne prend pas en compte le potentiel dont disposent les manuels ou la radio pour améliorer la performance des étudiants ayant des résultats en dessous de la performance optimum. Mais, l'évaluation essaye plutôt d'évaluer leur impact comme cela se ferait dans un pays type en développement" (Janison, 1982, p.559). Donc, simplement fournir des manuels ne suffit pas sans la conception d'une méthode pour s'assurer que les enseignants utilisent les manuels comme prévu.

L'évaluation a utilisé une méthode de sélection au hasard qui reposait sur une évaluation de projet qui existait déjà. Dans le cadre du projet Radio Nicaragua, un mécanisme d'ensemble avait permis la mise en place de procédures de sélection et de test au hasard afin d'évaluer l'efficacité du programme d'enseignement par la radio. Le projet existant avait déjà classé tous les établissements primaires dans trois provinces au Nicaragua dans les catégories *radio* ou *contrôle* en utilisant la procédure d'échantillonnage aléatoire structuré suivant le facteur de l'urbanisation (près de 30 pour cent des étudiants fréquentent des établissements situés en zones rurales, mais des effectifs aussi importants de classes ont été sélectionnés par rapport à chaque catégorie).

L'évaluation du manuel a exploité cette méthode qui existait déjà en sélectionnant les établissements expérimentaux et de contrôle de la façon suivante : d'abord, les évaluateurs ont obtenu une liste de tous les établissements avec les classes éligibles pour chacune des six catégories (trois provinces, rurales ou urbaines). Ensuite, ils ont désigné les établissements au hasard pour l'expérimentation et le contrôle à partir de leurs listes officielles par rapport à chaque catégorie, et les établissements ont été alors utilisés dans l'ordre de la liste (un établissement qui avait refusé de participer a été remplacé par l'établissement suivant sur la liste). Des demandes de participation de classes appartenant aux groupes de contrôle ont été rejetées, et toute utilisation du matériel d'expérimentation avait été contrôlée par les auteurs. Il est utile de noter que dès le départ, la méthode d'évaluation a fait état de l'éventualité de ce problème d'ordre politique. L'équipe d'évaluation a clairement défini ses objectifs dès le départ; l'équipe a obtenu l'accord et l'appui officiel à sa démarche, et l'équipe a également préparé des procédures claires et systématiques pour le programme.

L'étude a ainsi sélectionné au hasard 88 classes: 48 radios et 40 établissements de contrôle. Vingt établissements parmi les établissements de contrôle ont reçu des manuels pour chaque enfant, et les enseignants ont des instructions écrites et orales ainsi que les versions des tests destinées aux enseignants. La composante radio comprenait 150 leçons quotidiennes de mathématiques combinées à des fiches de travail pour étudiants ainsi que des instructions écrites et orales à l'intention des enseignants.

Une décision intéressante qui avait été prise était l'absence délibérée de supervision des groupes de contrôle. Cela a été vraiment difficile parce que l'absence de supervision a rendu difficile l'évaluation de l'utilisation. Cependant, l'effort consenti du point de vue influence de comportement a été jugé trop important. Des visites surprises qui avaient été acceptées comme solution de compromis n'ont pas pu être faites en raison des troubles politiques pendant l'année d'évaluation et ont été faites l'année suivante.

Une seconde décision a été de faire passer les tests par le personnel du projet au lieu des enseignants des classes eux-mêmes. Cela a vraiment augmenté les coûts administratifs mais a réduit les erreurs potentielles inhérentes aux tests. Les étudiants ont subi un pré-test de préparation en mathématiques pendant les trois premières semaines de cours. Le post-test qui avait pour objectif d'évaluer la performance des étudiants était prévu au cours des trois dernières semaines de l'année mais il a été administré deux semaines plutôt à cause de problèmes politiques. Les étudiants avaient autant que possible les mêmes conditions pour les deux tests lors de leur administration parce que les tests avaient la même durée de temps et les instructions avaient été enregistrées sur magnétophone.

III. Données

Il y a deux leçons principales à tirer de la composante collecte de données. La première est que les difficultés logistiques sont souvent inévitables. Malgré sa bonne conception, il y a eu un ensemble de problèmes pour la mise en oeuvre d'un jeu parfait de pré-tests et de post-tests de comparaison. Bien qu'il y ait eu un total de 20 classes de contrôle, 20 classes de manuels, et 47 classes de radio, le nombre de notes de pré-test et de post-test était différent dans chaque groupe à cause d'inscriptions tardives, d'absences, et d'incapacité de passer les tests en raison « d'effectifs pléthoriques ». Les informations personnelles sur les étudiants ne semblent pas avoir été collectées.

La deuxième leçon à tirer, c'est l'imagination dont on fait preuve les évaluateurs pour concevoir le post-test afin de réduire les charges tout en obtenant les informations nécessaires. Il y avait un certain nombre de problèmes à surmonter:

- Il n'y avait pas au Nicaragua de tests standards en utilisation.
- Le test était supposé évaluer la performance par rapport aux objectifs du programme.
- Le test était supposé refléter la performance par rapport à chaque thème pour faciliter une évaluation de l'efficacité de l'intervention sur chaque thème ainsi que sur l'ensemble.

Les évaluateurs ont utilisé une méthode d'échantillonnage à matrices multiples pour faire face à ces problèmes. Le test avait deux types de questions: celles posées à tous les étudiants dans la classe (40 éléments G) et celles posées à des sous-groupes d'étudiants (44 éléments D). Tous les éléments I avaient été testés dans chaque classe; un quart de tous les éléments G avaient été testés dans chaque classe. Cela permet aux chercheurs de procéder au hasard à la répartition des unités sur deux dimensions: les établissements et les types de tests. La moyenne des résultats post-test pour le groupe expérimental et le groupe de contrôle est obtenue en ajoutant la moyenne des résultats pour chaque test, et en calculant les erreurs types sur la base de la variance résiduelle après avoir isolé les principaux effets relatifs aux éléments et aux étudiants.

Les informations relatives à l'utilisation des manuels ont été également collectées au cours de l'année qui a suivi l'intervention dans 19 des 20 établissements qui utilisent les manuels.

IV. **Techniques économétriques**

L'organisation de l'évaluation a montré qu'une simple comparaison des moyennes des groupes expérimentaux et de contrôle serait appropriée, et c'est ce qui a été fait. L'approche peut être délicate s'il y a des divisions multiples et des interventions multiples, ce qui a été le cas pour cette évaluation. Donc, les évaluateurs ont également utilisé une approche simple de régression. Ici, la classe servait d'unité d'analyse, et la moyenne des résultats de la classe au post-test a régressé par rapport à la moyenne des résultats du pré-test et des fictifs pour les interventions de la radio et des manuels, un fictif urbain-rural, et le résultat moyen de la classe au pré-test comme variables indépendantes.

Un élément important dans toute évaluation c'est, de savoir si les différents groupes sont affectés différemment par le même traitement. Cela peut souvent être réalisé, comme cela a été fait au cours de cette évaluation, par l'utilisation de variables interactives. Les différences entre les zones rurales et urbaines peuvent être prises en compte par interaction entre le fictif urbain-rural et l'intervention. La différence au niveau de l'effet de l'intervention basé sur les résultats des premiers tests a été obtenue par interaction des résultats des premiers tests avec l'intervention.

V. **Ceux qui l'ont exécuté**

La Banque Mondiale a appuyé le projet de recherche, mais il était intégré au Projet conjoint Radio Mathématiques de l'Agence Américaine pour le Développement International et du Ministère de l'Education Nationale du Nicaragua.

VI. **Résultats**

Les auteurs ont trouvé que les manuels et la radio, comme variables, avaient eu des impacts significatifs sur les performances des étudiants: la disponibilité des manuels a permis aux étudiants d'améliorer leurs résultats au post-test de 3,5 éléments corrects, alors que pour les leçons à la radio l'amélioration a été de 14,9 éléments – résultats assez substantiels étant donné que la déviation type de la classe est de 8,3 et que celle des éléments individuels est de 11,8. Les leçons à la radio et les manuels ont tous les deux été plus efficaces dans les établissements ruraux et pourraient éventuellement jouer un grand rôle dans la réduction des différences entre la qualité urbaine et rurale. Ces résultats semblent être indépendants du niveau de performance initial de la classe, comme l'indiquent les résultats du pré-test. Les auteurs attribuent la différence de résultats entre les interventions à la radio et celles des manuels à des différences relevant de l'utilisation des manuels, surtout lorsque les enseignants sont mal formés.

VII Leçons tirées

Il y a trois leçons principales à tirer: l'importance du politique dans les décisions de conception, l'utilité de méthodes imaginatives de tests, et les difficultés liées au travail sur le terrain. D'abord, l'économie politique de méthodes aléatoires a été mise en valeur dans cette étude: il y a sans aucun doute de fortes pressions politiques qu'on peut être amené à supporter et qui méritent d'être prises en compte tôt et avec l'appui du gouvernement. Deuxièmement, les auteurs ont pu, grâce à la conception imaginative de tests, évaluer les différents aspects des résultats d'apprentissage sans avoir à administrer des tests qui sont en fait de longs tests. Enfin, les évaluateurs ont clairement soulevé un certain nombre de questions liées au travail sur le terrain: comment superviser l'adoption effective de manuels et qui devrait administrer les tests.

VIII. Source

Jamison, Dean T., Barbara Serle, Klaus Galda, and Satephen P. Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement." *Journal of Educational Psychology* 73 (4): 556-67.

Annexe 1.13 : Impact des programmes alternatifs de recouvrement des coûts sur l'accès aux soins de santé et l'équité au Niger

I. Introduction

Description du projet. La capacité de recouvrer une partie des coûts des soins de santé est primordiale dans la fourniture des soins de santé. On sait peu de choses, cependant, sur l'impact des différentes stratégies sur la qualité des soins et du bien-être. L'évaluation fait l'estimation de l'impact sur la demande de soins de santé de deux programmes pilotes de recouvrement de coûts dans le secteur (non hospitalier) de soins de santé primaire au Niger. Le Niger est un pays économiquement pauvre et rural ; les coûts de santé publique représentent 5 à 6 pour cent du budget gouvernemental et le gros de ce financement est mal dirigé vers les hôpitaux et le personnel. Le gouvernement voulait évaluer les conséquences des différents mécanismes de paiement et en a choisi deux : un mécanisme de financement de frais pour service et un autre pour service majoré de taxe ; tous deux combinés avec l'amélioration de la qualité et de la gestion. Le gouvernement voulait particulièrement voir comment la nécessité pour les soins de santé pouvait apporter des changements, notamment chez les groupes vulnérables, et si de telles améliorations étaient durables.

Points forts de l'évaluation. Les différents mécanismes de paiement ont été exécutés dans trois districts, à raison d'un par traitement et par contrôle. L'évaluation a utilisé une conception quasi-expérimentale basée sur les enquêtes sur les ménages combinées aux données administratives sur les coûts d'utilisation et de gestion. L'évaluation est particulièrement intéressante parce qu'elle s'adresse directement aux questions économiques au moyen d'un instrument d'enquête qui demande aux interlocuteurs leur désir de payer pour le service amélioré. Cette reconnaissance explicite du fait que certains résultats ne sont pas, en eux-mêmes, suffisants pour garantir un projet durable est une contribution extrêmement importante. Un autre aspect utile est l'évaluation explicite de l'impact de l'intervention sur les différents groupes cibles (les enfants, les femmes, les villages sans structures sanitaires et les citoyens les plus démunis).

II. Thèmes de recherche et méthode d'évaluation

Les questions principales concernaient l'impact du traitement sur (a) la nécessité pour et l'utilisation des structures de soins de santé primaire, (b) les groupes cibles spécifiques (les démunis, les femmes et les enfants), (c) l'accès financier et géographique, (d) l'utilisation de services de substitution, et (e) la pérennité des améliorations dans le cadre des coûts de recouvrement (coûts pour les patients et médicaments ainsi que le revenu et la volonté de payer).

Les trois districts de santé ont été sélectionnés au niveau de trois provinces à partir d'un registre administratif. Bien que tous trois fussent identiques en termes de caractéristiques économiques, démographiques et sociales, ils étaient ethniquement différents. Chaque district disposait d'un centre de santé, avec un volet de santé maternelle et infantile, un poste de santé et d'un médecin ainsi que de dispensaires ruraux.

Quatre améliorations de qualité et de gestion ont été introduites dans les deux districts de traitement ; aucune n'a été exécutée au niveau des deux districts de contrôle. En particulier, les stocks initiaux de médicaments avaient été livrés ; le personnel formé dans le diagnostic et le traitement ; un système de gestion de stock de médicaments et de gestion financière a été installé et le personnel formé pour son utilisation ; la capacité de supervision fut augmentée pour renforcer la gestion.

Les deux mécanismes de fixation de prix ont été introduits en même temps. Le premier représentait les frais par épisode, avec des frais de 200FCFA (0,66\$US) par utilisateur âgé de plus de cinq ans, des frais de 100FCFA (0,33\$ US) pour un utilisateur âgé de moins de cinq ans. Le deuxième

combinait une taxe annuelle de 200FCFA payée par les contribuables et des frais de 50FCFA par utilisateur âgé de plus de cinq ans et 25FCFA pour les enfants de moins de cinq ans. Le revenu annuel était inférieur à 300\$US par habitant. Chaque programme offrait des exonérations pour les groupes cibles. Les fonds étaient gérés au niveau du district.

III. Données

Les trois districts ont été choisis à partir de données administratives. Des enquêtes sur les ménages ont été effectuées, dont l'une a été une enquête référentielle alors que les deux autres ont été combinées avec des archives et des données administratives. Chaque enquête a collecté des informations individuelles et sur les ménages à partir d'un échantillon de 1800 ménages choisis au hasard. L'enquête référentielle disposait de renseignements sur 2833 individus qui ont été malades au cours des deux semaines antérieures à l'enquête et 1770 femmes enceintes ; l'enquête finale disposait de données sur 2710 individus malades et 1615 femmes enceintes. Les données administratives comprenaient des renseignements détaillés sur les dépenses mensuelles, sur la consommation et l'administration de médicaments, l'entretien du personnel et les reçus de frais avec l'utilisation des structures sanitaires. Les renseignements ont été collectés au cours de l'année précédant l'intervention, l'année de départ, (mai 1992 - avril 1993) et l'année après l'intervention.

IV. Techniques économétriques

L'étude combine des comparaisons de moyennes avec des techniques logiques simples, les dernières étant utilisées pour représenter les changements d'utilisation. En particulier, la réponse individuelle est de savoir si la structure de santé a été utilisée (PI) pour spécifier le modèle suivant :

$$\text{Logit (Pt)} = X\beta + * (A + B).$$

Ce modèle qui vérifie le vecteur des caractéristiques individuelles X ainsi que des variables fictives A et B , a été comparé avec

$$\text{Logit (Pi)} = X\beta + *a A + *b B.$$

Les variables fictives A et B sont définies de manière différente. Dans la première batterie des régressions, A se rapporte à la période durant le traitement, B à la période avant le traitement et les régressions sont traitées par sous-groupe (les groupes cibles spécifiques) et par district. Dans la seconde batterie des régressions, A et B sont utilisés pour former six paires semblables de comparaisons de chaque district avec un autre district au cours du traitement. Dans chaque cas, les auteurs vérifient si $(+a + +b) = *$. Les effets de l'accessibilité géographique et financière sont représentés dans la matrice X par les mesures de distance, de temps de marche et de revenus, respectivement. La catégorie omise n'est pas précisée pour chaque cas, à partir de la discussion. Il n'est pas non plus clair si les erreurs normales des estimations ont été corrigées pour la nature imbriquée de la conception de l'échantillon.

Bien que les techniques de logit soient un moyen efficace de réponse à trois ou quatre questions de recherche - utilisation des procédures, effet sur les sous-groupes et effets de l'accessibilité géographique et financière - la quatrième question, l'effet des changements dans le recouvrement des coûts, est analysée par le biais de données administratives et de moyens simples de comparaisons. Une préoccupation claire dans cette dernière approche, qui n'était pas clairement soulevée, concerne la possibilité de l'existence de parti pris dans la phase suivant le report des résultats. En particulier, il existe une chance morale si les administrateurs sont évalués sur la base de la réponse réussie du traitement.

Les effets du traitement sur les systèmes de santé de substitution ont été étudiés par le biais de techniques économétriques décrites auparavant.

V. **Ceux qui l'ont exécuté**

Le Ministère de la Santé Publique a entrepris l'enquête avec le soutien financier et technique de l'Agence Américaine pour le Développement International et de la Banque Mondiale. L'évaluation proprement dite fut effectuée par Francis Diop, Abode Yazbeck et Ricardo Bitran de Abt Associates.

VI. **Résultats**

L'étude a révélé que le système de taxe plus frais a produit plus de revenu par habitant que le système de taxe sans frais, en plus d'être beaucoup plus populaire. Le système basé sur la taxe a également donné de meilleurs résultats en termes d'opportunités d'accès aux soins de santé améliorés au profit des démunis, des femmes et des enfants. Cependant, vu que la géographie est une barrière importante dans l'accès aux soins de santé, le système de taxe redistribue de manière plus efficace le coût des soins de santé des personnes les plus proches des structures de santé au profit des personnes qui en sont les plus éloignées.

Le district qui a exécuté les frais pour service a connu une chute sur le nombre de visites initiales mais une croissance dans la recherche des services de soins de santé – en comparaison avec la croissance dramatique dans les deux districts ayant appliqué la taxe plus les frais. Ce fait important pourrait être attribué à la croissance de la qualité des services associée à la qualité des améliorations, qui augmentent le plus souvent le coût.

L'élément coût – en particulier celui des médicaments – associé à la qualité et à la réforme de la gestion s'est également avéré effectif et durable. Le recouvrement des frais majorés des taxes au niveau des districts a avoisiné et dépassé les 100 pour cent mais était substantiellement plus bas que celui au niveau des districts pratiquant les frais pour service. De plus, il y avait une plus grande volonté de payer au niveau des premiers que des derniers.

Le résultat principal est que l'approche taxe plus frais est à la fois plus efficace dans la réalisation des objectifs et a plus de succès auprès des populations. Cependant, l'évaluation a démontré également que le manque d'accessibilité géographique aux structures de santé constitue une barrière importante à leur utilisation. Ceci suggère qu'il existe des questions de distribution associées au passage au système des frais plus taxes : les ménages qui vivent loin des structures de santé auraient tendance à financer ceux vivant à proximité de ces centres.

VII. **Leçons tirées**

Il y a beaucoup de leçons à tirer de cette évaluation. La première est la voie à contours multiples dans laquelle elle évalue l'impact du projet sur des dimensions multiples liées à la durabilité: non seulement au plan du recouvrement des coûts mais également sur la qualité et la réaction des groupes cibles affectés. La deuxième est l'attention accordée au détail dans la collecte de données concernant tant les instruments administratifs que l'enquête, qui ont ainsi produit les fruits à travers la capacité d'identifier exactement les éléments de l'intervention qui ont réussi et pourquoi. Enfin, l'analyse de l'impact sur chaque groupe cible s'est avérée être particulièrement utile dans les recommandations de politique.

VIII. **Sources**

Diop, F. A Yazbeck, et R. Britan. 1995. "L'Impact des Programmes de Substitution de Recouvrement des Coûts sur Accès et l'Equité au Niger." *Politique de Santé et de Planification* 10 (3): 223-40.

Wouters, A. 1995. "Improving Quality through Cost Recovery in Niger." 10 (3) :257-70.

Wouters, A. 1995. "Amélioration de la Qualité à Travers le Recouvrement des Coûts au Niger". 10 (3): 257-70.

Annexe 1.14 : Résultats scolaires dans les établissements primaires aux Philippines : Evaluation de l'impact à partir de quatre expériences

I. Introduction

Description du projet. Dans la plupart des pays en développement, les taux élevés de déperdition scolaire et l'apprentissage inadapté dans l'enseignement primaire constituent une préoccupation pour les décideurs. Ceci est certainement le cas aux Philippines: près du quart des enfants philippins abandonnent l'école avant d'avoir terminé la sixième année, et ceux qui abandonnent ont appris moins de la moitié de ce qui leur a été enseigné. Le Gouvernement s'est engagé dans un Programme d'Intervention contre la Déperdition Scolaire (DIP) en 1990-1992 pour faire face à ces problèmes. Quatre expériences avaient été entreprises: fourniture de matériel d'apprentissage de niveaux multiples (MLM), repas à l'école (SL), et chacune de ces interventions est combinée avec un partenariat parent-enseignant (PTP). La première approche permet aux enseignants de gérer les enseignements selon les besoins de chaque étudiant et coûte beaucoup moins cher que les repas scolaires. Les partenariats parent-enseignant ne coûtent presque rien mais peuvent contribuer à aider les étudiants à apprendre aussi bien à la maison qu'à l'école.

Points forts de l'évaluation. L'évaluation est valable en ce sens qu'elle cherche de façon explicite à développer des capacités au niveau du pays hôte afin que l'évaluation devienne une partie intégrante des nouvelles initiatives, et que les préalables relatifs aux données soient connus avant plutôt qu'après la mise en oeuvre des projets à venir. Mais il y a des problèmes qui surgissent, et l'évaluation est très claire sur les attentes. Une autre contribution importante a trait à la vérification de la solidité des résultats par des approches économétriques différentes. Enfin, l'analyse de la rentabilité appliquée à la fin est importante en ce sens qu'elle reconnaît explicitement que des résultats significatifs ne suffisent pas: des interventions peu coûteuses peuvent encore être meilleures relativement à des interventions coûteuses.

II. Thèmes de recherche et méthode d'évaluation

La question principale a trait à l'évaluation de l'impact de quatre interventions différentes sur la déperdition et les performances des étudiants. Cependant, la méthode d'évaluation est conditionnée par des considérations aussi bien pratiques que programmatiques. L'équipe DIP a suivi un processus de sélection scolaire en trois étapes:

- Deux districts dans chacune des cinq régions du pays ont été identifiés comme municipalité à faible revenu. Dans l'un des districts les traitements étaient choisis pour constituer un ensemble d'éléments de contrôle, MLM, ou MLM-PTP; dans l'autre de contrôle, le traitement SL, ou SL-PTP avait été utilisé. La décision des deux types d'interventions était prise par tirage au sort.
- Dans chaque district l'équipe a sélectionné trois établissements qui (a) avaient tous les niveaux d'instruction, et une classe par niveau; (b) avaient un niveau élevé de déperdition scolaire; et (c) n'avaient pas de programme de nutrition en place.
- Les trois établissements dans chaque district étaient sélectionnés soit comme établissements de contrôle soit pour l'application de l'une des interventions sur la base d'un tirage au sort.

Chaque intervention a été assignée au hasard à toutes les classes au niveau de cinq établissements, et les pré-tests et post-tests ont été administrés en 1991 et 1992 à toutes les classes dans les 20 établissements ainsi que dans 10 établissements de contrôle.

III. Données

La procédure de collecte de données est instructive en elle-même. La collecte de données de référence fut commencée en 1990-1991, et les interventions ont été exécutées en 1991-1992. Les informations détaillées ont été collectées relativement à 29 établissements, quelque 180 enseignants, et près de 4,000 étudiants au cours de chacune des deux années. Bien que les questionnaires aient été très détaillés, cela s'est avéré inutile: seule une petite partie de ces informations a été effectivement utilisée, ce qui suggère qu'une partie de l'effort du processus d'évaluation puisse être utilement réduite. Les pré-tests et post-tests ont été administrés au début et à la fin de chaque année académique dans trois matières: Mathématiques, Langue philippine et Langue anglaise.

Les données ont été structurées de façon longitudinale pour les étudiants et les établissements. Malheureusement, les identifications sur les étudiants ont fini par ne pas être les mêmes pour les étudiants et les établissements dans la période entre les deux années. Il est utile de mentionner que cela n'était pas connu a priori et n'est devenu évident que lorsque six mois de travail ont permis de découvrir des irrégularités internes. Il n'a pas été possible de reprendre les identificateurs d'origine avec le Département de l'Education des Philippines. Heureusement, il avait été possible de sauver les données pour les premières années, ce qui a permis de faire quelques analyses longitudinales.

IV. Techniques économétriques

La structure de la procédure d'échantillonnage a soulevé un certain nombre de problèmes économétriques intéressants: un type de problèmes relatifs à la déperdition scolaire et un type de problèmes relatifs aux résultats du test. Dans chaque cas, il y a deux types de contrôles évidents: l'un étant le groupe de contrôle des établissements, et l'autre étant l'enquête de référence effectuée l'année avant l'intervention. Les auteurs les ont pris en charge différemment.

Dans l'analyse de la déperdition, il est normal d'effectuer une différence dans l'approche et de comparer le changement dans le taux moyen de déperdition au niveau de chaque classe d'intervention pendant la période entre les deux années avec le changement dans le taux moyen de déperdition pour les classes de contrôle. Mais deux questions se sont posées immédiatement. D'abord, bien que les résultats aient été très importants du point de vue taille, ils n'étaient significatifs que pour l'intervention MLM, peut-être en raison de la petite taille de l'échantillon. Cela n'est pas étrange pour ce type de procédure et peut devenir endémique étant donné le manque de financement pour les études de grande échelle dans le contexte d'un pays en voie de développement. Deuxièmement, une vérification brève pour savoir si les caractéristiques et performances des étudiants ont été réellement les mêmes dans tous les établissements au cours de l'année qui a précédé les interventions a révélé qu'il existait des différences significatives de caractéristiques. Ces deux facteurs ont amené les auteurs à vérifier la solidité de leurs résultats à travers les techniques logistiques de régression pour le contrôle des caractéristiques personnelles (PC) et de l'historique de la famille (FB). Le résultat principal n'a pas changé. Mais la technique de régression a révélé un facteur indirect de déperdition mais important néanmoins, à savoir les mauvais résultats scolaires. Ceci a naturellement amené au second type d'analyses qui a mis l'accent sur la performance scolaire.

Un autre type de préoccupations d'ordre économétrique a été soulevé au cours de l'évaluation de l'impact de l'intervention INTER sur la performance académique d'individus I en classe S à un moment t (APist) que les auteurs ont formulé comme suit :

$$AP_{ist} = \delta_0 + \delta_1 AP_{ist-1} + \delta_2 PC_1 + \delta_3 FB_1 + \delta_4 LE_{st} + \delta_5 CC_1 + \delta_6 INTER_{jt} + \varepsilon$$

où LE correspond au contexte de l'apprentissage et CC aux conditions des classes. L'un des premiers problèmes c'est d'expliquer la corrélation groupée en erreurs qui est possible chez les étudiants qui sont dans les mêmes classes et les mêmes établissements. Le deuxième problème c'est d'essayer de saisir l'aspect hétérogène non observé. Le troisième problème associé c'est celui de la distorsion de la sélection. Le premier problème a été traité en appliquant une correction Huber-White à toutes les erreurs types. Le deuxième problème pourrait, en principe, être analysé à un niveau individuel en

utilisant les différences de résultats de tests comme variable indépendante. Mais les auteurs soutiennent que cela est inapproprié parce qu'ils présupposent que la valeur de δ_1 c'est 1, ce qui n'a pas été mis en évidence par les tests. Ils ont donc retenu la caractéristique de la variable dépendante décalée, mais cela a entraîné un autre problème – l'une des erreurs endogènes de régression. Ce problème est traité en instrumentalisant le résultat du pré-test de chaque matière avec les résultats du pré-test des autres matières. Les auteurs constatent, cependant, que la réduction des erreurs engendre des risques – une réduction dans l'efficacité – et ils présentent alors les résultats à la fois des variables de racines et instrumentales les moindres. Les auteurs utilisent les variables constantes des établissements et des enseignants pour contrôler les aspects hétérogènes non observés dans LE et CC.

Le troisième problème est également un problème endémique à la littérature et pour lequel il n'y a pas de solution complètement acceptée: distorsion de la sélection. En clair, en raison du fait qu'il y a des différences de taux de déperdition, la performance académique individuelle est sujette à la condition de ne pas abandonner l'école. Bien que ce problème ait été discuté dans le cadre de la procédure Heckman à deux étapes, il y a une insatisfaction prononcée envers cette procédure pour trois raisons: sa sensibilité à l'assomption de la distribution normale, le choix et la pertinence des variables appropriées à utiliser lors de la première étape, et sa dépendance fréquente de l'identification au travers de la non linéarité de sa première étape. Malheureusement, il n'y a pas encore de consensus sur une alternative appropriée. Celle qui a été proposée par Krueger assigne aux étudiants abandonnant l'école leur classement à l'issue du pré-test et les ramène à la régression. Ainsi, les auteurs rapportent trois types de résultats: la simple régression des résultats, l'approche Krueger, et la procédure Heckman.

V. **Ceux qui l'ont exécuté**

La collecte de données a été réalisée par le Bureau de l'Enseignement Primaire du Département Philippin de l'Education, de la Culture, et des Sports. L'analyse a été réalisée par un employé de la Banque Mondiale et deux chercheurs académiciens.

VI. **Résultats**

L'étude évalue l'impact de ces interventions sur la déperdition scolaire de la sixième à la première et sur les résultats aux tests en première année utilisant une approche de différence à différences, des techniques instrumentales variables, et la méthode de sélection Heckman. L'impact de matériels à niveaux multiples -- particulièrement avec le partenariat parent-enseignant -- sur la déperdition et l'amélioration de la performance académique dépend de différents facteurs et également d'une certaine rentabilité. L'impact des repas à l'école a été en général faible. Une composante intéressante pour l'étude a été une analyse de rentabilité – qui est arrivée à la pertinente conclusion que tout ne s'arrête pas aux bons résultats! En particulier, un calcul direct des coûts directs et indirects du programme amène à la conclusion que l'approche MLM est à la fois efficace et rentable.

Dire que les programmes de nutrition scolaire sont inefficaces est peut être exagéré, mais il est possible qu'une approche mieux ciblée pour ces programmes soit appropriée. De plus, en raison du fait que la période entre la mise en oeuvre et l'évaluation du programme est trop courte, l'évaluation ne peut pas estimer les effets à long terme des interventions.

VII. **Leçons tirées**

Plusieurs leçons ont été tirées de cette procédure d'évaluation. La principale leçon a été qu'il y a souvent beaucoup de déperditions lorsque l'on va trop dans les détails – une quantité trop importante d'informations longitudinales peut se perdre si les informations pertinentes telles que l'unicité des identifications sur une période de temps sont perdues. La deuxième leçon c'est que très peu d'informations collectées au cours d'enquêtes détaillées ont été utilisées et qu'on aurait pu éviter aux interviewés les efforts substantiels consentis inutilement. Troisièmement, l'étude met l'accent sur la valeur des différentes approches économétriques et les avantages de retrouver la systématité à travers les techniques. Quatrièmement, cette étude est

exemplaire dans son utilisation de l'analyse de la rentabilité – identifiant et valorisant les coûts des différentes interventions. Finalement, bien que des erreurs aient été clairement faites au cours de l'étude, les auteurs constatent que la raison d'être principale de cette étude était de développer la capacité d'évaluation aux Philippines. Le fait que le DIP ait été amélioré et évalué signifie que cette capacité peut être entretenue au sein des ministères de l'éducation.

VIII. **Source**

Tan, J. P., J. Lane, and G. Lassibille. 1999. "Schooling Outcomes in Philippine Elementary Schools: Evaluation of the Impact of Four Experiments." *World Bank Review*, September.

Annexe 1.15 : Evaluation de l'impact des projets de routes rurales sur la pauvreté au Vietnam

I. Introduction

Description du projet. Les pistes rurales sont largement soutenues par la Banque Mondiale et les autres bailleurs de fonds comme instrument de réduction de la pauvreté. Le Projet I du Transport Rural au Vietnam a été lancé en 1997 avec un financement de la Banque Mondiale pour une mise en oeuvre sur trois à cinq ans. L'objectif du projet est d'améliorer le niveau de vie dans les zones pauvres en réhabilitant les routes et ponts existants et en développant l'accès au marché. Dans chaque province participante, des projets sont identifiés pour la réhabilitation à travers les critères de moindres coûts (taille de la population qui en bénéficiera et coût du projet). Mais, dans un effort d'améliorer le ciblage de la pauvreté, 20 pour cent des fonds de chaque province peuvent être mis à côté pour les zones montagneuses à faibles densités peuplées par des minorités ethniques où les projets ne seraient pas éligibles dans le cadre des critères de moindres coûts.

Evaluation de l'impact. En dépit d'un consensus général sur l'importance des routes rurales, il est étonnant qu'il n'y ait que peu d'évidence pratique sur la taille et la nature des bénéfices émanant d'une telle infrastructure. L'objectif de l'Evaluation d'Impact des Routes Rurales au Vietnam c'est de déterminer comment le bien-être des ménages change dans les communes où il y a des interventions de projets routiers par rapport aux communes qui n'ont pas de tels projets. Le problème principal de l'évaluation c'est de réussir à isoler l'impact de la route de la multitude d'autres facteurs qui sont en train de changer aujourd'hui dans le milieu rural Vietnamien en raison de la transition en cours vers une économie de marché.

L'évaluation a commencé simultanément avec la préparation du projet, vers le début de 1997, et est toujours en cours. Les résultats ne sont pas encore disponibles. L'évaluation est contraignante en ce sens qu'elle constitue l'une des premières tentatives d'évaluation de l'impact d'un projet de routes rurales sur le bien-être – c'est à dire l'essentiel même en terme de détermination si les projets luttent réellement contre la pauvreté. La méthode vise à améliorer les efforts antérieurs de l'infrastructure d'évaluation en combinant les éléments suivants: (a) collecte de données de référence et de suivi d'enquête, (b) inclusion de contrôles adéquats afin que les résultats soient robustes à des facteurs invisibles qui influencent aussi bien le programme de placement que les résultats, et (c) suivi assez soutenu du projet (grâce à des séries de collectes de données) pour saisir son impact total sur le bien-être.

II. Méthode d'évaluation

La méthode d'Evaluation de l'Impact des Routes Rurales au Vietnam met l'accent sur des données de référence (avant intervention) et de suivi (après intervention) d'enquête pour un échantillon de communes avec projets et sans projets. Les contrôles appropriés peuvent être identifiés à partir des communautés sans projets grâce à des techniques de comparaison appariée. Les données de référence permettent de faire des comparaisons avant et après ("réflexives") des indicateurs de bien-être au niveau des communautés de projets et des communautés de groupes de contrôle. En théorie, le groupe de contrôle sélectionné par la technique de comparaison appariée, est identique au groupe du projet suivant les caractéristiques visibles, et non visibles de telle sorte que les résultats au niveau des programmes communautaires peuvent être attribués à l'intervention du projet.

III. Collecte de données et techniques d'analyse

Les données collectées pour les besoins de l'évaluation comprennent des enquêtes effectuées au niveau des communes et des ménages, avec des données de base provenant des districts,

provinces et projets. Les enquêtes de référence et de suivi au niveau des communes et des ménages avaient été effectuées entre 1997 et 1999, alors que la deuxième série et la troisième série d'enquêtes effectuées avec un intervalle de deux années sont planifiées. L'échantillon de l'enquête comprend 100 communes avec projets et 100 communes sans projet, situées dans 6 des 18 provinces couvertes par le projet. Les communes avec projets ont été sélectionnées au hasard à partir de listes de toutes les communes avec des propositions de projets dans chaque province. Une liste avait alors été tirée à partir de toutes les autres communes restantes dans les districts ayant des propositions de projets, et à partir de cette liste, des communes de contrôle ont été sélectionnées au hasard (L'idéal c'est que les groupes de contrôle ne diffèrent des groupes avec projets que s'ils ne reçoivent pas d'interventions. Et pour des raisons logistiques, il était souhaitable de limiter le travail sur le terrain à certaines régions. Les contrôles étaient donc choisis dans le voisinage des districts, et en réalité dans les mêmes districts que les communes où le traitement est administré. Les districts sont grands et l'influence d'une commune avec projet sur une commune sans projet est donc improbable, mais cela doit être vérifié attentivement.) Les techniques de comparaison par score de tendance basées sur les caractéristiques des communes seront utilisées pour tester la sélection des contrôles, et tout contrôle avec des attributs inhabituels ayant rapport avec les communes du projet sera retiré de l'échantillon. Un modèle logit de participation des communes au projet sera estimé et utilisé pour s'assurer que les communes de contrôle ont les mêmes scores de tendance (valeurs anticipées par le modèle logit).

La base de données de la commune se fonde sur les données administratives existantes collectées annuellement par les communes et traitant de questions démographiques, d'utilisation des terres, et d'activités de production; de plus il y a également des données provenant d'enquêtes effectuées au niveau des communes dans le cadre de l'évaluation. L'enquête couvre des caractéristiques générales, les infrastructures, l'emploi, les moyens de subsistance, l'agriculture, la terre et les autres biens, l'éducation, les soins de santé, les programmes de développement, les organisations communautaires, les finances de la commune, et les prix. Ces données seront utilisées pour élaborer un nombre d'indicateurs de bien-être au niveau de la commune et pour tester les impacts des programmes dans le temps.

L'objectif principal de l'enquête sur les ménages c'est d'obtenir des informations sur l'accès des ménages aux différentes facilités et services et de voir comment cela évolue dans le temps. Le questionnaire destiné aux ménages avait été administré de façon aléatoire à 15 ménages dans chaque commune, et couvrait les thèmes comme l'emploi, les biens, les activités de production et d'emploi, l'éducation, la santé, la commercialisation, le crédit, les activités communautaires, l'accès à la sécurité sociale et aux programmes de pauvreté, et le transport. En raison des capacités limitées d'enquêtes à l'intérieur, il n'y a pas eu de tentative de collecte de données complexes nécessaires à la mise en place d'un indicateur de bien-être au niveau des ménages (comme le revenu ou la consommation). Cependant, un certain nombre de questions avait été inclus dans l'enquête qui reprend des questions posées lors de l'enquête sur les niveaux de vie au Vietnam. En utilisant cette information et tant d'autres encore sur les caractéristiques des ménages, qui sont communes aux deux enquêtes, des techniques de régression seront utilisées pour estimer la position de chaque ménage dans la répartition du bien-être au niveau national. Une base de données au niveau du district a également été préparée pour mettre les données de la commune en contexte, y compris les données sur la population, l'utilisation de la terre, l'économie, et les indicateurs sociaux. Chacune de ces enquêtes doit être répétée suivant le calendrier de la commune.

Les informations existantes ont été utilisées pour mettre en place deux autres bases de données. Une importante base de données a été mise en place au niveau de la province pour aider à comprendre la sélection des provinces dans le projet. Cette base de données concerne toutes les provinces du Vietnam et couvre un éventail plus grand de variables socio-économiques. Finalement, une base de données au niveau de chaque projet pour chacune des zones de projet enquêtées a été également mise en place afin de contrôler l'ampleur du projet et sa méthode de mise en oeuvre dans l'évaluation de l'impact pour le projet.

Les données de référence seront utilisées pour modéliser la sélection des sites de projet en mettant l'accent sur les processus économiques, sociaux et politiques sous-jacents. Les étapes ultérieures seront utilisées pour comprendre les gains quantifiables au niveau de la commune, sur la base de la sélection. L'approche analytique aura un rôle de "double différenciation" avec les méthodes appariées. L'appariement sera utilisé pour sélectionner le groupe idéal de contrôle parmi les 100

communes sans projet. Les résultats au niveau des communes avec projet seront comparés à ceux obtenus par les communes de contrôle, aussi bien avant qu'après l'introduction des projets de routes. L'impact du programme est ainsi identifié comme étant la différence entre les résultats dans les zones de projet, antérieurement et postérieurement à l'intervention du projet, moins la différence de résultats correspondante dans les zones de contrôle appariées. Cette méthodologie permet d'avoir une estimation correcte des impacts de projets en présence de facteurs invisibles invariables dans le temps qui influencent à la fois la sélection des zones de projet et les résultats. Les résultats seront mis en évidence par le fait que les séries de données sont riches en indicateurs de performance et en variables explicatives. Les indicateurs de performance à examiner comprennent les rendements agricoles au niveau de la commune, la diversification des sources de revenus, les opportunités d'emploi, l'utilisation et la distribution des terres, la disponibilité des produits, les services et les facilités, le bien être sanitaire et la distribution.

IV. Coûts d'évaluation et administration

Coûts. Le coût total de l'évaluation à ce jour est de 222,500 \$, ou 3.6 pour cent du coût total du projet. Ce montant comprend 202,500 \$ couvrant les deux premières étapes de la collecte de données et une subvention de recherche d'un montant de 20,000 \$. Le temps de travail et les frais de voyages du personnel de la Banque Mondiale ne sont pas inclus dans ces coûts.

Administration. L'évaluation a été élaborée par Dominique van de Walle, membre du personnel de la Banque Mondiale. Un consultant indépendant, économiste et chercheur dans le domaine de la pauvreté rurale, avait été recruté comme superviseur de l'étude au niveau du pays. Ce consultant a recruté et formé les superviseurs d'équipes, organisé toute la logistique, et supervisé toutes les données.

V. Source

Van de Walle, Dominique. 1999. *Evaluation de l'Impact sur la pauvreté des projets de Routes Rurales*. World Bank, Washington, D.C. Processed.

¹ utiliser un processus non-aléatoire dans la sélection d'un groupe de contrôle qui correspond le plus exactement aux caractéristiques des bénéficiaires du programme

Ndt. 1 acre = moins de 40 ares.

³ Sélection aléatoire des bénéficiaires du programme au sein d'un groupe éligible

Annexe 2

Termes de Référence

Exemple I : Projet de nutrition et de développement de la petite enfance de l'Ouganda

Termes de référence pour un Cabinet de consultation en vue de sa contribution à l'évaluation de Projet

I. Généralités

Le gouvernement de l'Ouganda a souscrit à un crédit de l'Association pour le Développement International (IDA) en faveur d'un projet de nutrition et de développement de la petite enfance. Le projet se concentre essentiellement sur l'amélioration de la qualité de vie des enfants au-dessous de six ans et le développement de la capacité des familles et des communautés à s'occuper des enfants. Le projet vise spécialement le développement de la petite enfance à travers l'amélioration de la nutrition, de la santé, de l'état psychosocial et cognitif des enfants au-dessous de six ans en Ouganda.

II. Raison fondamentale pour investir dans le développement de la petite enfance

Investir dans le développement de la petite enfance (DPE) comporte des avantages réels non seulement pour les enfants et les parents mais également pour des communautés entières et le pays. Une croissance physique et un développement mental rapides se produisent chez le nouveau-né et au cours de la petite enfance ; à deux ans, le cerveau d'un enfant est presque entièrement développé. Les capacités intellectuelles sont également largement développées à l'âge de quatre ans. Une croissance et un développement physique et mental adéquats pendant la petite enfance renforcent la préparation pour l'école, améliorent le maintien des enfants à l'école et contribuent à la confiance du capital humain. Les enfants des milieux défavorisés en particulier, peuvent bénéficier de l'assistance accordée à la petite enfance, comblant de ce fait les fossés et les inégalités liés à la pauvreté. Une bonne santé et une bonne nutrition sont aussi importantes que la stimulation mentale, lorsque l'enfant doit développer des structures conceptuelles solides plus tard. La synergie entre la nutrition, la santé, et la stimulation mentale est si importante que les effets positifs tangibles sur la croissance et le développement de l'enfant ne puissent être réalisés qu'à travers une approche intégrée.

III. Objectifs et stratégies du projet

L'objectif de développement du projet est d'améliorer la croissance et le développement des enfants au-dessous de six ans dans les aspects de nutrition, de santé, psychosociaux et intellectuels. La réalisation de ces objectifs à la fin de la période des cinq années de mise en place du projet sera évaluée grâce aux repères suivants : (a) la réduction du taux d'enfants chétifs dans le préscolaire d'un tiers par rapport aux taux établis en 1995 dans les zones de projet ; (b) la réduction de la présence du rachitisme dès l'entrée des enfants dans les écoles primaires, d'un quart par rapport aux taux établis en 1995 dans les zones de projet ; (c) l'amélioration du développement psychosocial et intellectuel des enfants ; (d) la réduction des taux de redoublement et d'abandon au cours des premières années des écoles primaires, et (e) le développement des qualités d'esprit d'entreprise et de potentialité économique des mères et de tous ceux qui s'occupent des enfants.

Le projet soutient le Programme National Ougandais de l'Action en faveur des Enfants et le Plan d'Action pour l' Eradication de la Pauvreté. Le projet améliore en particulier l'aptitude scolaire des jeunes enfants, contribuant ainsi à la réalisation de l'objectif de l'enseignement primaire universel. La stratégie principale du projet est d'augmenter la capacité des familles et des communautés à mieux s'occuper des enfants en âge préscolaire (zéro à six ans) à travers l'amélioration de leur connaissance sur la croissance et le développement des enfants, leur capacité à élever des enfants, la nutrition et les soins de santé et des activités génératrices de revenus pour les femmes.

IV. Approche du projet

Le projet est un programme basé sur un processus et donnant la priorité à la zone concernée plutôt que de simples propositions. Les ressources doivent être mises en place progressivement au sein des communautés à la suite d'un processus de planification participative pour en assurer l'appropriation et la pérennité. Le programme impliquera une collaboration entre le gouvernement et les entités non gouvernementales, y compris les organisations non gouvernementales locales et internationales (ONG), et les communautés. En tant que programme multisectoriel englobant la santé, la nutrition, l'enseignement pré-scolaire, l'assistance à l'enfance, la génération de l'épargne et de revenus, l'approche consistera à relier divers services gouvernementaux et entités non gouvernementales pour créer un service global orienté vers le développement des enfants. Le projet soutiendra une variété d'options - un menu de programme - en rapport avec les besoins des enfants du niveau pré-scolaire et de leurs familles.

V. Composantes du projet

Composante 1 du projet – Interventions communautaires intégrées d'assistance à l'Enfance. Cette composante soutient les objectifs du gouvernement : (a) d'améliorer l'éveil de conscience parentale sur les aspects principaux d'assistance à l'enfance, de croissance et de développement à travers l'éducation parentale, le contrôle et la promotion de la croissance de l'enfant, la formation et la sensibilisation ; et (b) de donner la possibilité aux communautés de soutenir des programmes de développement de l'enfant par la création de compétence, par l'acquisition des capacités pour générer des revenus et par des subventions de soutien. L'objectif est de réduire la malnutrition (poids inférieur par rapport à l'âge) des enfants d'un tiers à la fin de la période des cinq années dans les zones de projet et améliorer la préparation des enfants pour l'instruction à l'école primaire, contribuant de ce fait à la campagne pour rendre l'enseignement primaire universel. Le plan du gouvernement est de couvrir éventuellement toutes les zones ; cependant, les interventions dans cette phase seront mises en œuvre dans 25 zones choisies par le gouvernement, sur la base du degré de malnutrition, de mortalité infantile et du taux d'inscription au niveau du primaire. Le projet comprend les interventions suivantes, reliées entre elles :

(a) Education parentale. Cette sous composante augmentera la compréhension des parents et des soignants au sujet des aspects principaux de l'assistance à l'enfance, de la croissance, et du

développement, y compris la nutrition de l'enfant, la santé, et le développement intellectuel et psychosocial. Les qualités y relatives seront renforcées chez les parents. Le développement des qualifications et de la connaissance parentale améliorera en retour la santé, le développement psychosocial et le bien être des enfants et finalement, leur réceptivité à l'éducation au niveau primaire. Le programme mobilisera des groupes de mères de famille (et de parents) au niveau de la communauté avec le support de matériels de projet en langues locales, d'un contrôle technique et d'un système de communication. Des matériels didactiques simplifiés pour adultes avec un niveau d'instruction très bas ont été testés avec succès en Ouganda. L'accent sera mis sur le renforcement des pratiques en matière d'assistance à l'enfance qui favorisent la croissance et le développement appropriés de l'enfant y compris la nutrition et la santé de l'enfant (allaitement exclusif au sein et pratiques appropriées de sevrage : en particulier la période de l'introduction de s aliments de sevrage, aussi bien que les types de nourriture donnés et la préparation de la nourriture, la promotion de la croissance de l'enfant, et l'utilisation des vermifuges), le développement psychosocial, la stimulation de la connaissance et le soutien social, et l'hygiène et l'amélioration des pratiques de santé à domicile. Les interventions ci-dessus seront renforcées et soutenues par une activité plus étendue (la Journée de l'enfant) organisée au niveau paroissial pour permettre aux communautés d'accéder à un certain nombre de services concernant l'enfant au moyen d'achats dans un centre commercial. Une étude de l'impact de l'approvisionnement en albendazole d'anathelminth aux enfants en bas âge dans des paroisses sélectionnées sera également entreprise au cours des journées paroissiales de l'enfance et mesurera l'effet des traitements effectués tous les six mois sur le gain de poids.

(b) Création de la compétence de la communauté et habileté à l'assistance à l'enfance. Cette sous-composante comporte deux activités en corrélation : (a) le développement de la capacité de la communauté, réalisé à travers des ateliers de planification et sensibilisation de la communauté, et (b) la formation à l'esprit d'entreprise pour augmenter les revenus des mères de familles et des assistants sociaux.

Composante 2 du projet – Subventions de soutien à la communauté pour le développement de l'enfant. Deux types de subventions seraient à la disposition des communautés :

(a) Les subventions de soutien de la communauté – subventions aux communautés offertes sur la base de contributions similaires des communautés. Ces subventions et contributions des communautés couvriront des activités conçues pour soutenir les interventions pour le développement de l'enfant qui font partie des lignes directrices et de la liste des activités contenues dans le manuel d'exécution du projet. Pour être qualifiées pour cette subvention, les communautés devront fournir les contributions de contre –partie qui peuvent être sous forme de marchandises, de travaux ou de services. Les exemples d'utilisation de telles subventions sont la construction et le fonctionnement des centres communautaires d'assistance à l'enfance, des centres d'assistance à domicile à l'enfance ou la production et la vente des aliments de sevrage. Les composantes des subventions de soutien seront mises en application dans les mêmes 25 zones incluses dans la composante 1.

(b) Les subventions d'innovation – subventions rendues disponibles aux communautés pour traiter les problèmes relatifs aux enfants. La subvention d'innovation facilitera l'application des interventions en dehors de la liste des interventions décrites par les subventions de soutien à la communauté (a) ci-dessus. Comme le terme l'indique, les fonds d'"innovation" seront employés pour soutenir les communautés à différents niveaux dans la mise en application "d'idées innovatrices" en vue de l'amélioration du niveau de vie des enfants au sein de leurs communautés. Les subventions d'innovation seront accessibles aux communautés de la même manière que dans le cas des subventions de soutien aux communautés : c'est-à-dire, des propositions seront préparées par les communautés suivant un exercice de planification participative, seront ensuite examinées par un sous-comité et envoyées pour le financement par le projet.

Composante 3 du projet – Programme national de soutien pour le développement de l'enfant. Cette composante se compose des activités centrales du programme et des initiatives de politiques conçues

pour soutenir les programmes de zones dans les composantes 1 et 2 et pour fournir la garantie de la qualité pour les activités de première ligne du projet au niveau de la communauté. Cette composante inclut : (a) le contrôle du programme et son évaluation ; (b) le soutien pour la prévention des insuffisances des éléments micro-nutritifs ; (c) le développement du programme de la petite enfance ; (d) la formation des formateurs pour la petite enfance ; et (e) l'information, l'éducation, et la communication.

VI. Les mesures de mise en application

La mise en application du projet relève de la responsabilité du gouvernement de l'Ouganda assisté par des organisations non-gouvernementales dans le cadre de la décentralisation et le transfert des pouvoirs à des niveaux inférieurs comme stipulé dans les politiques nationales. La communauté (LC-1) est l'unité d'opération pour l'exécution des services, bien que la structure de coordination implique également les niveaux de la paroisse (LC-2), du sous-comité (LC-3), et de la zone (LC-5).

En plus, le projet compte utiliser la sensibilisation et les consultations des acteurs, la mobilisation de la communauté, la planification participative de la communauté, le renforcement de la capacité pour la pérennité à tous les niveaux, tout en renforçant la complémentarité avec des structures et programmes nationaux existants. Les structures politiques et institutionnelles existantes devraient être utilisées de façon multisectorielle. La transparence et la responsabilisation devraient également être assurées à tous les niveaux.

VII. Coordination du projet

Niveau national. Un comité de pilotage du projet composé des ministères concernés, des donateurs et des représentants des ONG sera responsable de la conduite générale de la mise en œuvre du projet.

Le comité qui sera dirigé par le secrétaire permanent au Ministère des Finances, de la Planification, et du Développement Economique, guidera le projet pour les aspects politiques, examinera et approuvera la qualité et l'efficacité de la mise en œuvre. Le comité de coordination du projet fera également des suggestions pour améliorer le budget annuel de la zone et les plans de travail du projet. Un petit bureau de coordination du projet (PCO), composé d'un coordinateur, d'un adjoint, d'un comptable qualifié et d'un petit nombre de personnel de soutien sera basé dans le secteur des services sociaux du Ministère des Finances, de la Planification, et du Développement Economique et aura la responsabilité de la coordination des activités journalières sur le plan national.

Niveau de la zone. Un comité de coordination multisectoriel existant—le comité de coordination de Zone (DCC)—sera identifié par le Chef de l'Administration pour prendre la responsabilité de coordonner le projet au niveau de la zone. L'Administrateur identifiera une personne centrale parmi les fonctionnaires du gouvernement qui coordonnera les activités en relation avec les ONG et les autres activités dans le cadre du projet. L'ONG principale sera incluse en tant que membre du Comité de Coordination de Zone. Dans les zones où aucune ONG, ayant la capacité et la base de compétences adéquates ou appropriées, ne peut être identifiée ou renforcée pour émerger comme ONG principale, la mise en œuvre se fera à travers l'administration de la zone.

Niveau de sous-zone. Un comité sectoriel existant semblable à celui du niveau de la zone sera responsable de coordonner les activités du projet au niveau du sous-comité. Ce sous-comité de coordination facilitera également les liaisons entre les structures existantes et celles du projet et, ensemble avec l'ONG principale pour la zone approuveront les plans de travail annuels et les demandes des ONG/CBO (Organisations communautaires) pour le projet dans le sous-comité.

VIII. Evaluation de l'impact du projet

Le gouvernement de l'Ouganda voudrait évaluer divers aspects du projet afin de s'assurer de son efficacité et guider la conception d'autres projets nationaux de nutrition et de développement de la petite enfance. De plus, comme la Banque Mondiale compte sur ce projet pour informer potentiellement d'autres pays sur les services du projet de développement de la petite enfance et de la nutrition, elle a inclus le projet dans l'évaluation des programmes de développement de la petite enfance dans trois pays et fournira une assistance technique sur la base de subvention au bureau de coordination du projet (PCO) pour aider les activités de recherche spécifique. En particulier, deux études d'évaluation de l'impact des interventions spécifiques du projet seront entreprises en tant qu'éléments du projet global.

- Etude de la journée paroissiale de santé de l'enfant pour évaluer la couverture des traitements anthelminthiques proposés au niveau paroissial au cours des journées de santé de l'enfant et leur impact sur le gain de poids des enfants de moins de six ans par l'utilisation d'une conception expérimentale randomisée.
- Recherche par sondage en utilisant une référence, un réexamen pour évaluer (a) l'impact des traitements anthelminthiques et des activités globales du projet sur le développement intellectuel, la santé, et la nutrition des enfants de moins de six ans ; (b) l'impact de la composante de l'éducation des soignants et de la campagne de communication des mass médias sur la connaissance, l'attitude, et les systèmes d'encadrement des enfants des principaux soignants; et (c) l'impact de la formation à la base en gestion, aux activités génératrices de revenus et la formation d'un groupe d'épargne et de crédit, et l'octroi de subventions communautaires pour le bien-être des ménages et de la communauté.

Le bureau d'étude choisi fournira le soutien technique et logistique pour les études citées ci-dessus et sera invité à prendre part, en tant que membre du groupe de recherche local chargé de la mise en œuvre, de la conception, de la collecte des données et de l'analyse nécessaire pour achever les deux études d'évaluation d'impact. Ce bureau sera le partenaire principal du bureau de coordination du projet, des chercheurs locaux et des chercheurs de la Banque Mondiale et de l'Université d'Oxford qui entreprendront l'évaluation de l'impact.

IX. Vue d'ensemble des études

Etude n°1 : Impact du traitement des vers au cours des Journées de l'Enfant de la paroisse. Un certain nombre d'études font état de l'impact du traitement des enfants en âge d'aller à l'école avec de l'anthelminthique. Cependant, il n'existe qu'un seul test sur un procédé aléatoire effectué à grande échelle qui montre un grand effet sur le gain de poids pour les enfants en âge préscolaire. Ceci a soulevé la question de savoir si un tel impact pouvait être réalisé avec les enfants africains. Ainsi, le projet de développement de la petite enfance et de la nutrition inclura une étude aléatoire de l'impact de la fourniture de vermifuge l'albendazole d'anthelminth, aux enfants à bas âge dans 25 paroisses sélectionnées au cours des Journées de l'Enfant des paroisses et mesurera l'effet des traitements sur le poids des enfants tous les six mois. Des données seront recueillies dans ces paroisses de même que dans les 25 groupes de contrôle qui organiseront également des journées de santé infantile mais n'administreront pas l'albendazole systématiquement. Dans le cas où les traitements à l'anthelminthique sont administrés avec succès et que des effets bénéfiques sont notés sur les enfants d'Ouganda, le programme du traitement à l'anthelminthique pourrait alors être recommandé pour toutes les autres régions.

Puisqu'il s'agit d'un essai scientifique sous contrôle, la sélection des paroisses qui seront invitées à administrer l'albendazole sera effectuée par le bureau de coordination du projet et ceci à partir d'une liste de paroisses où des journées de l'enfant seront organisées (cette liste sera fournie par les ONG

opérants dans les régions). Le bureau de coordination du projet choisira également les paroisses qui serviront de groupe de contrôle. Cette conception expérimentale est essentielle pour le succès de l'évaluation.

Le cabinet de consultation s'assurera que les ONG locales responsables de l'organisation des Journées de santé infantile dans les paroisses comprennent les raisons d'une telle conception expérimentale et qu'elles se conforment à la stratégie. Chaque enfant âgé de 12 mois ou plus et chaque enfant de moins de 6 ans ayant participé à cette expérience organisée dans les 25 paroisses indiquées recevra un comprimé de 400 mg à croquer d'albendazole breveté. L'albendazole sera administré tous les six mois. Dans le cas où les ONG choisiraient d'organiser des journées de l'enfance plus fréquemment, l'anthelmintique serait toujours administré tous les six mois et pas plus souvent.

Les enfants dans les paroisses où l'albendazole est administré aussi bien que les enfants des 25 paroisses de contrôle désignées seront pesés à chaque journée de l'enfance et leurs poids seront enregistrés sur leur propre carnet de santé et sur le registre de la communauté. Les enfants qui sont trop petits pour se tenir sans aide sur la balance seront pesés dans les bras de leur mère après que la balance a été réglée sur zéro avec la mère se tenant toute seule sur la balance.

Ces poids seront enregistrés au dixième le plus proche (0,1) d'un kilogramme. Les données figurant sur les registres de la communauté sont la responsabilité des ONG locales malgré le fait que le bureau travaillera avec les ONG pour s'assurer que le système de collecte de données est compatible avec les objectifs de l'étude.

Le consultant transcrira ces poids sur un proforma à concevoir en collaboration avec les conseillers techniques de la Banque Mondiale et de l'Université d'Oxford. Cette transcription de données sera entreprise semestriellement après la journée de l'enfance dans la zone du projet. En plus de l'identification de l'enfant (une seule combinaison de l'identification de la paroisse, de l'identification du village et de l'identification individuelle enregistrée à la fois sur la carte de l'enfant et le registre de la communauté), les données sur le proforma incluront le sexe de l'enfant ; la date de naissance de l'enfant relevée sur sa carte de santé ou, si l'information n'est pas disponible, de l'âge de l'enfant relevé sur le registre de la paroisse ; la date de la kermesse des enfants où le poids a été enregistré; et si oui ou non l'enfant a pris une dose d'albendazole. Ces données seront informatisées à Kampala. L'identification individuelle fournira la base pour fusionner les données de différentes périodes ; ainsi, l'identification doit être enregistrée chaque fois que les données seront transcrites et doivent rester constantes pour un enfant pendant toute la durée du projet.

Les circonstances et les conditions locales à chaque journée de l'enfance qui peuvent dissuader les mères d'être présentes seront également enregistrées. Celles-ci incluront des données sur l'état des récoltes et les conditions atmosphériques qui peuvent toutes les deux décourager des mères d'être présentes. Toutes les méthodes et occasions spéciales utilisées pour annoncer chaque journée de l'enfance seront enregistrées parce que les différentes formes de publicité peuvent influencer sur le taux de participation. Les données enregistrées devraient également inclure une évaluation du nombre d'enfants venant d'autres paroisses qui ont visité la journée de l'enfance et qui n'ont pas obtenu des numéros d'identification de la part des organisateurs de cette journée.

L'expérience s'étendra sur deux années. Ainsi, le consultant enregistrera les données cinq fois pour chaque paroisse. C'est-à-dire que le consultant rassemblera les données au début du projet et à 6, 12, 18, et 24 mois après le déclenchement du projet.

Une copie complète des données sera envoyée semestriellement au bureau de coordination du projet. Ces copies seront considérées comme étant les services pouvant être proposés pour la première étude du projet. Une analyse préliminaire sera entreprise à l'Université d'Oxford deux fois par an. Cependant, le bureau d'étude devra nommer un représentant qui participera à l'analyse principale à entreprendre à la fin des deux années. Ce représentant bénéficiera d'un fonds de voyage et d'hébergement afin de travailler sur l'analyse à Oxford. Les fonds pour ce voyage seront budgétisés dans une ligne séparée d'éléments et n'auront donc pas besoin d'être inclus dans le contrat couvert par la demande des propositions.

Etude N° 2 : Impact global des interventions du projet de nutrition et de développement de la petite enfance. Les enquêtes sur les ménages et sur la communauté constitueront les informations de base et de suivi requises pour évaluer l'impact des diverses activités du projet. Les enquêtes auront plusieurs modules qui évalueront :

- Le développement et la croissance cognitifs des enfants de moins de six ans résultant des traitements anthelminthiques et des activités globales du projet – L'étude n°2 évaluera la croissance longitudinale et les résultats psychosociaux et cognitifs de développement sur un groupe d'enfants dans les communautés participant au projet (avec ou sans traitement anthelminthique) comparé à un groupe d'enfants dans les communautés ne participant pas au projet. Les deux groupes seront suivis pendant deux années ou plus. Cette étude complètera donc celle sur l'administration de vermifuge effectuée au niveau des paroisses et permettra d'avoir une meilleure compréhension de la décision d'emmenager les enfants aux journées de l'enfance et d'évaluer si, avec le temps, la participation entraîne une augmentation dans les mesures de développement cognitif. De plus, en incluant les communautés qui ne bénéficient d'aucun service de développement de la petite enfance, l'étude évaluera si l'ensemble des services mis en place conduit à une amélioration du mode alimentaire et du développement cognitif.
- Les changements opérés dans la connaissance, l'attitude et les soins prodigués aux enfants par les soignants résultant de l'éducation parentale du projet et la campagne faite par les mass médias.
- L'amélioration de la santé et de la nutrition des enfants de moins de six ans résultant des contrôles de croissance, de l'éducation en santé préventive et en nutrition, des traitements anthelminthiques et des activités globales du projet.
- Le bien-être des ménages résultant des subventions communautaires de la formation, la formation à la base en gestion, des activités génératrices de revenus et de la formation de groupe d'épargne de crédit.
- Les caractéristiques de la communauté et les changements résultant des interventions du projet (ou non) qui pourraient avoir un impact sur le bien-être de l'enfant pendant la durée du projet.

Sélection d'échantillon. Cette étude sera basée sur les données d'une enquête de base rassemblées au moment où les services étaient administrés la première fois aux communautés et sur une enquête de suivi effectuée auprès de ces mêmes ménages deux ans après l'étude initiale. Un tiers de l'échantillon sera constitué à partir des mêmes 25 paroisses qui font partie du groupe sous traitement anthelminthique et un tiers constitué des groupes de contrôle étudiés dans l'étude n° 1. En outre, un tiers de l'échantillon viendra des villages de 25 paroisses dans les mêmes régions que les groupes de traitement, mais qui ne sont pas prévus pour bénéficier des services du projet de développement de la petite enfance et de la nutrition. Trente ménages seront sélectionnés dans chaque paroisse. Ceci implique 750 ménages par couche sociale. A cause de la déperdition dans l'échantillon prévue (2.250 au total) dans l'enquête initiale, 5 à 10 % en moins de ménages sont attendus dans la nouvelle enquête à effectuer.

Pour constituer le groupe d'échantillon dans les paroisses de traitement et de contrôle, une liste de tous les ménages visés (il y a approximativement 700 ménages dans une paroisse en moyenne) sera établie, probablement par un résident de la communauté. Sur cette liste figureront le nom du chef de famille, une indication de l'emplacement du ménage et du nombre d'enfants de moins de six ans au sein du ménage. Cette liste répond à deux objectifs. D'abord, un échantillon de 30 ménages comprenant au moins un enfant de moins de six ans par paroisse sera sélectionné au hasard. En second lieu, le nombre total d'enfants de moins de six ans servira comme estimation de la couverture potentielle des enfants au cours des journées de l'enfance, aidant ainsi dans la détermination du taux de participation.

Etant donné que le projet de développement de la petite enfance et de la nutrition aura moins de contact avec les communautés qui n'ont aucune activité de ce projet, la sélection des ménages ne bénéficiant pas du programme de développement de la petite enfance devrait utiliser un échantillonnage de groupe pour réduire les coûts de l'établissement d'échantillon. En particulier, un sous-district qui n'est pas concerné par le projet devrait être sélectionné pour chaque sous-district faisant partie du groupe de

traitement, de préférence celui qui lui est contigu. Une liste de toutes les paroisses de ces sous-districts devra être établie et un tirage au sort de 25 paroisses du nombre total de paroisses sera opéré. Deux villages de chaque paroisse sélectionnée seront alors choisis, toujours en utilisant la liste de tous les villages de la paroisse. Cette démarche réduit le nombre de villages où un recensement est nécessaire. Le recensement—semblable à celui effectué dans les paroisses de traitement et de contrôle—constituera la liste des ménages utilisée pour tirer les échantillons de 30 ménages par paroisse. Ceci sera la troisième couche de l'enquête.

L'enquête initiale de base devra être entreprise vers la mi-1999. Ce calendrier est basé sur la nécessité de connaître les sous-districts et les paroisses dans lesquelles les ONG vont opérer afin d'utiliser le procédé d'échantillonnage suggéré. Ce calendrier est aussi basé sur l'idée selon laquelle la sélection et la formation des coordonnateurs des ONG ne seront pas achevées avant fin 1998.

Cependant, l'élaboration du questionnaire et son test devraient être entrepris beaucoup plutôt (que début 1999) afin que l'enquête puisse démarrer aussitôt que les ONG auront identifié les paroisses dans lesquelles elles vont opérer. Comme les données de base devraient être disponibles avant la première séance d'administration de vermifuge, la période idéale pour l'enquête de base coïncide avec l'organisation initiale de la communauté qui aboutira à la journée de l'enfance. Parce que l'échantillon de 30 familles dans chaque paroisse est petit par rapport à la population totale, il est probable que la collecte de données de l'enquête ne perturbe pas d'autres activités ou surcharge les communautés de travail. La collecte de données dans les groupes de contrôle (ceux qui ont des activités des ONG mais n'administrent pas de vermifuge et ceux avec ni l'un ni l'autre) devrait être faite simultanément avec la collecte de données dans le groupe de traitement.

Instruments d'enquête. Les questionnaires de base à utiliser pour les besoins de l'enquête sont des questionnaires sur les ménages (qui rassemblent les données au niveau des ménages et des individus) et les questionnaires sur les communautés.

X. ENQUETE DE MENAGES

Les données sur les ménages seront collectées en utilisant un programme pré-codé. Ce programme sera établi sur le modèle des enquêtes sur le niveau de vie utilisé dans plus de 30 pays. Une première ébauche sera élaborée par des chercheurs de la Banque Mondiale. Cependant, l'instrument sera abrégé pour i) s'accommoder aux besoins particuliers du projet et, ii) s'adapter aux conditions locales en utilisant des groupes de centres d'intérêt et un procédé de pré-test entrepris par le bureau d'étude. Le questionnaire sur les ménages contiendra des modules pour rassembler des données sur les points suivants :

1. Les caractéristiques socio-démographiques : une liste des personnes résidant dans le ménage au cours des 12 derniers mois, leur âge et leur sexe, de même que leur niveau d'instruction et leur type d'emploi (s'il y en a). La forme de codage devra mentionner les parents de tous les enfants, s'ils sont présents—au cas contraire, mentionner si les parents sont toujours en vie. Une liste détaillée des avoirs sera élaborée pour servir d'indicateur sur le niveau socio-économique.
2. La connaissance, l'attitude et les pratiques : le questionnaire rassemblera également les informations sur la connaissance, les attitudes et les soins prodigués aux enfants par les principaux assistants sociaux.
3. Les données anthropométriques : les poids seront enregistrés au dixième près (0,1) d'un kilogramme pour tous les enfants de moins de six ans en utilisant des balances à affichage numérique qui seront fournies. En outre, les tailles seront enregistrées pour tous les enfants âgés de 2 à 6 ans. Le système pré-test sera utilisé pour déterminer s'il est possible d'enregistrer aussi les poids des mères de ces enfants (au cas où elles vivent dans les ménages).
4. L'évaluation cognitive : le bureau d'étude travaillera avec d'autres consultants locaux ou internationaux du bureau de coordination du projet pour intégrer les tests du développement cognitif de l'enfant dans la collecte de données globale de la zone. Dans l'enquête de base, un test

de développement cognitif reconnu sur le plan international sera appliqué aux enfants âgés de 4 à 5,99 ans (moins de 6 ans). Ce test sera également appliqué à la même catégorie d'âge dans la deuxième partie de l'étude, permettant une comparaison des groupes. En outre, un sous-groupe d'enfants âgé de 6 à 7,99 ans au moment de la deuxième partie recevra aussi ce test. (Le tableau 2.1 en annexe récapitule cette stratégie.)

En outre, les évaluations de la connaissance basées sur le contenu spécifique du programme et une évaluation dynamique peuvent être développées au cours de la deuxième partie de l'étude. L'inclusion de ces mesures sera évaluée pendant le déroulement du projet. En conclusion, une mesure de performance scolaire sera développée pour évaluer la connaissance acquise au cours de la première année scolaire et sera administrée à un sous-groupe d'enfants plus âgés au cours d'une nouvelle enquête. Les tests déjà existants pourraient être adaptés.

5. La santé infantile: les données sur la morbidité (y compris le nombre et le type de symptômes, les degrés de gravité, la durée), les modes d'accès et d'utilisation des services de santé, d'hygiène et autres.
6. L'économie des ménages : la meilleure approche pour acquérir cette information sera explorée intensivement dans la phase de pré-test et sera évaluée ensemble avec des conseillers du bureau de coordination du projet avant la finalisation du questionnaire. Les variables peuvent inclure les dépenses alimentaires ; les activités agropastorales; la consommation des produits de la maison ; les dépenses non alimentaires ; les caractéristiques de logement ; les inventaires des biens durables ; l'emploi ; les activités économiques ; les revenus; la terre ; les cultures et les animaux ; les revenus des activités du projet ; les entreprises familiales; la possession de biens ; l'information de crédit et d'épargne sur la somme d'argent et de biens prêtés et empruntés, si de l'argent et de biens ont été empruntés au cours des 12 derniers mois ; l'épargne et la dette nette au jour de l'interview ; l'information sur les prêts y compris les traites et les raisons de l'emprunt, et le nombre de prêts venant de la même source ; et la localisation de l'épargne, s'il en existe, y compris la banque, la caisse d'épargne logement, la caisse d'épargne rurale, et autres. Cette information fera partie seulement des enquêtes de base et du projet.

XI. Enquête Communautaire

Les questionnaires communautaires seront utilisés pour recueillir des informations sur les conditions locales qui sont communes à tous les ménages dans le secteur. La meilleure approche pour procéder à la collecte de ces informations sera explorée intensivement dans la phase de pré-test et sera évaluée ensemble avec les conseillers du bureau de coordination du projet avant de finaliser le questionnaire. Les variables peuvent inclure :

1. L'information démographique : le nombre de ménages, la population totale, le nombre d'enfants de moins de six ans, les groupes ethniques et les religions ;
2. L'information économique, y compris les activités économiques principales et les schémas de migration pour le travail;
3. L'infrastructure : l'accès aux routes, l'électricité, l'eau courante, le marché, la banque et le transport en commun. Les conditions de l'infrastructure locale, telles que les routes, les points d'approvisionnement de carburant et d'eau, l'électricité et les moyens de communication ;
4. Les conditions et pratiques agricoles locales : le type de culture pratiquée dans la communauté, le rythme des plantations et des récoltes, le mode général de vente de la récolte et des données qualitatives sur les niveaux de précipitation, les conditions climatiques et les saisons ;
5. L'éducation : le nombre et les types d'écoles maternelles, les organisations officielles et non officielles de programme de développement de la petite enfance, la distance aux écoles, le nombre de classes, les taux d'inscription, (en général et par sexe), le taux de fréquentation, le rythme de passage, les soins de santé et de nutrition fournis à l'école (par exemple, les programmes de santé scolaire, les repas scolaires) ;

6. La santé : le type d'infrastructure de santé et la distance à parcourir et le temps à mettre pour atteindre la plus proche de chacune des infrastructures de santé (hôpital, pharmacie, poste de santé, et autres). La distance et le temps de voyage pour atteindre le plus proche de chacun des nombreux types de personnel de santé (médecin, infirmière, pharmacien, sage femme, ouvrier de santé communautaire, et autres) ; et
7. Autres : le nombre et le type d'ONG/Organisations communautaires locales actives, les autres projets ou interventions relatifs aux enfants (par exemple, les campagnes de vaccination du gouvernement) et d'autres projets de développement communautaire.

Personnel d'enquête recommandé. Le personnel chargé de l'enquête devrait être constitué

comme suit :

- personnel d'enquête central: composé du directeur de l'enquête, du directeur de zone, du directeur de la gestion des données et du personnel de saisie de données qui sera respectivement responsable de la supervision globale de la zone, de la coordination et du contrôle de la collecte des données, de la saisie des données et des activités de gestion des données.
- personnel d'enquête de zone: les enquêtes sur le terrain seront menées par des équipes composées d'un superviseur, de deux (ou trois) intervieweurs responsables du questionnaire principal et des mesures anthropométriques et un chauffeur. Un nombre semblable de spécialistes participeront à l'administration des tests de développement cognitif aux enfants et seront sélectionnés et formés en collaboration avec les experts locaux et internationaux.
- Coordonnateur pour l'essai aléatoire du test : le coordonnateur aidera au développement des instruments de collecte de données, à la formation des responsables des ONG locales pour l'organisation des journées de l'enfance dans les paroisses, sur la conception expérimentale, la collecte de données et la transcription de ces données. Il ou elle supervisera la saisie des données et la gestion de l'étude de l'ensemble des données, et participera à l'analyse principale à réaliser à la fin de l'étude.

Organisation du travail sur le terrain. Le Cabinet de consultation participera à l'ébauche des instruments de terrain avant l'enquête pilote et sera le principal responsable du pré-test. Après le pré-test, le questionnaire sera reformulé (en collaboration avec les chercheurs de la Banque Mondiale) et traduit ensuite dans les langues locales.

Le Cabinet travaillera en collaboration avec d'autres consultants locaux et internationaux de recherche sélectionnés par le bureau de coordination du projet pour intégrer des tests de développement cognitif de l'enfant dans la collecte globale des données de terrain. Le chercheur local du programme de développement de la petite enfance, aidé par les conseillers internationaux choisira et adaptera le test cognitif principal à utiliser et formera ceux qui seront chargés d'administrer les essais.

L'organisation suivante de travail sur le terrain est recommandée. Cela est basé sur une expérience internationale et conçue pour assurer un contrôle de qualité. Une certaine variation de cette approche pourrait être acceptée en collaboration avec les chercheurs de la Banque Mondiale en partant de l'expérience du Cabinet et des autres conseillers du bureau de coordination du projet et de l'information obtenue durant l'enquête pilote.

Le travail sur le terrain sera organisé en petites équipes constituées d'un superviseur, de deux ou trois intervieweurs responsables du questionnaire principal et des mesures anthropométriques et un nombre semblable de spécialistes chargés d'administrer les tests de développement cognitif des enfants. Ce personnel sera formé à Kampala par le consultant local en recherche du programme de développement de la petite enfance en collaboration avec des conseillers internationaux sur les tests psychologiques. La formation inclura une discussion sur les objectifs de la recherche, le passage en revue de chaque étape de l'interview, la pratique de la formation dans le bureau, un essai sur le terrain et une récapitulation des expériences après cette répétition.

Une fois que les équipes sont bien formées, elles peuvent être engagées pour toute la durée de l'enquête, si possible. Cependant, du fait que quelques membres du personnel peuvent se révéler inadaptés pendant le travail sur le terrain, il est recommandé de former quelques membres du personnel supplémentaires. Il n'est pas conseillé d'engager un personnel pour un travail de quelques jours seulement dans une paroisse et d'engager encore un nouveau personnel pour la paroisse suivante, pour éviter d'avoir un personnel inexpérimenté. Tout le personnel devra recevoir une nouvelle formation au début de la nouvelle enquête.

Au cours de l'administration du test cognitif, les enfants doivent, dans la mesure du possible, être seuls avec le chargé de l'interview. En aucun cas, une autre personne (adulte ou enfant) ne doit répondre aux questions posées à l'enfant. Cependant, au cours de la nouvelle enquête, le test pour le sous-groupe des enfants âgés de 8 ans peut être administré en groupe si cela s'avère convenable. Le superviseur sera responsable de s'assurer que ceux chargés de l'interview entreprennent l'enquête dans les ménages sélectionnés pour l'échantillonnage sans substitution et que tous les enfants dans les groupes d'âge appropriés reçoivent les tests de développement cognitif. En outre, le superviseur repassera en revue chaque questionnaire après le test (avant le départ de l'équipe dans une nouvelle paroisse) pour s'assurer qu'il n'y a pas des éléments qui manquent dans le questionnaire et que toute information apparemment incohérente a été vérifiée.

Le cabinet d'étude fera saisir toutes les données de l'enquête aussitôt que possible après la collecte. Les copies des données spécifiques sur les ménages et les enfants, et les barèmes d'évaluation ainsi que la documentation nécessaire pour accéder aux données nécessaires seront informatisés et fournis au bureau de coordination du projet à la fin de l'enquête de base. Les questionnaires originaux doivent être conservés par le cabinet d'étude parce qu'il sera nécessaire d'accéder à ces informations au cours de l'analyse.

Les données au niveau de l'enfant doivent contenir les codes précis d'identification qui peuvent correspondre aux codes d'enquête de ménage. Bien que les seuls codes des personnes et des ménages fournis au bureau de coordination du projet ne doivent pas comporter les noms des ménages ou leur emplacement exact, ces informations doivent être enregistrées par le bureau de manière à permettre qu'une autre visite à ces ménages puisse se faire à une date ultérieure. Puisqu'une étape dans l'analyse lie les personnes de la nouvelle enquête à leurs résultats de test de départ, il est essentiel que les codes des personnes et des ménages puissent être gardés constants au cours des trois enquêtes.

XII. Tâches spécifiques des spécialistes de l'enquête

Le cabinet d'étude participera aux activités suivantes en collaboration avec le bureau de coordination du projet, les chercheurs locaux, les chercheurs de la Banque Mondiale et de l'université d'Oxford et les ONG exécutantes :

- Révision des programmes de travail.
- Développement et adaptation des instruments de collecte des données et de la documentation annexe, y compris la liste du matériel, les questionnaires, les guides de codage, les manuels de l'intervieweur et du superviseur, le manuel des opérations, le manuel de saisie de données et les procédures de terrain.
- Révision et ébauche des divers projets de documents, de la disposition, de la traduction, de la rétro traduction et du test sur le terrain. La fourniture des tirages et des versions électroniques de toute la documentation au bureau de coordination du projet. La transmission des questionnaires aux chercheurs de la Banque Mondiale pour leur revue avant l'enquête pilote.
- Etablissement de listes d'habitation et mise à jour de la cartographie. Les responsabilités de la liste des ménages et des habitations dans chaque paroisse sélectionnée incluent l'obtention des cartes graphiques de base, la préparation des matériels de liste, le contact des fonctionnaires locaux pour les informer de la préparation de la liste, l'identification des limites, l'élaboration des plans,

l'établissement systématique de la liste des ménages, l'obtention d'information préliminaire sur les ménages y compris le nom du chef de famille, une indication de l'emplacement du ménage et du nombre d'enfants de moins de six ans du ménage; les procédures de documentation au moment de la constitution de l'échantillon, à la fin du travail sur le terrain et à la fin de la constitution du fichier de données.

- Préparation du cadre d'échantillonnage (avec le spécialiste d'échantillonnage), de la formation du personnel pour mettre en application l'échantillonnage préparé, la supervision de l'étape de mise en application pour assurer la qualité de l'échantillonnage sélectionné et l'établissement d'un rapport détaillé présentant les grandes lignes de toutes les étapes impliquées dans la conception et la mise en place de l'échantillonnage.
- En consultation avec la Banque Mondiale, participer à la détermination d'une stratégie appropriée pour identifier des groupes de comparaison (c'est-à-dire, les paroisses ne participant pas au projet).
- Sélection et formation de ceux qui travailleront sur le terrain. Cette activité se compose de tout le travail nécessaire pour développer des matériels et des manuels de formation pour toutes les personnes impliquées dans le travail sur le terrain. La formation sera exigée pour ceux qui sont chargés des interviews, les superviseurs des interventions, les superviseurs des équipes, le personnel de saisie de données et le personnel anthropométrique.
- Opération de terrain, y compris la logistique pour la collecte des données et l'obtention du consentement des ménages et des individus et la tenue d'un registre d'étude sur les ménages.
- Production des rapports sur l'état d'avancement: le cabinet d'étude préparera des rapports sur l'état d'avancement du travail sur le terrain (à intervalles de six mois) dont une copie sera envoyée au bureau de coordination du projet et à la Banque Mondiale. Le cabinet d'étude devrait également préparer une description de base de l'enquête. Cette description devrait inclure le contenu de l'enquête, le plan d'échantillonnage et sa mise en application et les techniques utilisées pour le travail sur le terrain. Un questionnaire complet et une documentation de base devraient être joints en annexe.
- Développement d'un programme de saisie de données en utilisant un logiciel qui pourrait vérifier les séries et la cohérence des données et produire des rapports indiquant les données manquantes, les données en dehors des gammes autorisées et les réponses incohérentes.
- Nettoyage de données, saisie de données, gestion de base de données et plans de tableau, y compris le développement du programme de saisie de données, de la formation d'agent de saisie, de contrôles de qualité de données, et de directives d'utilisation des données. De même, les questions ouvertes de programmation des données, le contrôle des données anthropométriques sur la base de tableaux standard de référence.
- Application de l'accord de politique d'utilisation de données : le cabinet d'enquête et les chercheurs impliqués dans le processus de collecte et d'analyse des données signeront un protocole d'accord avec le bureau de coordination du projet qui énoncera explicitement la politique relative aux questions telles que l'accès aux données, les utilisateurs visés, les procédures pour obtenir des copies de données et la documentation, les règles de publication et de droit d'auteur.
- Conduite de l'analyse de données : le cabinet mènera des analyses de données exploratoires (par exemple, les fréquences, les tableaux de pourcentage et les tableaux-croisés) des variétés principales d'enquêtes et de leurs corrélations. Le cabinet mènera une étude statistique moderne de modélisation des impacts après les 2^e et 3^e tours d'enquêtes pour déterminer le progrès global au niveau des indicateurs sociaux (par exemple, la nutrition, la santé, les revenus et le développement communautaire) et les facteurs qui expliquent les changements ou le manque de changements.
- Production des rapports d'analyse : Le cabinet rendra compte des résultats après les 2^e et 3^e tours des études basées sur les analyses des indicateurs sociaux et des variables. Le cabinet travaillera

en collaboration avec le bureau de coordination du projet et la Banque Mondiale sur l'étude des journées de santé infantile des paroisses et sur la collecte d'impacts sur le développement cognitif, mais ne sera pas responsable des rapports définitifs sur le résultat de ces études.

Les tâches spécifiques de l'enquête de communauté comprennent les points suivants :

- Travailler en collaboration avec les conseillers du bureau de coordination du projet dans l'élaboration du questionnaire de communautés et explorer intensivement dans la phase de l'enquête pilote la meilleure approche pour rassembler l'information ;
- Travailler étroitement avec les agences de mise en oeuvre (ONG principale et locale) dans la collecte des données de la communauté ;
- Contacter les fonctionnaires et les dirigeants de la communauté pour expliquer l'approche d'évaluation de l'impact du projet et pour obtenir le consentement communal pour les enquêtes de recherche et l'étude de la journée de santé infantile ;
- Interviewer les informateurs principaux et obtenir les cartes, les listes et autres archives de la communauté ;
- Obtenir les listes des infrastructures de santé et d'éducation (préscolaire et primaire), y compris l'emplacement géographique, le secteur et le type d'établissement (par exemple, privé ou public);
- Obtenir les informations sur la démographie de la communauté, y compris le nombre de ménages et la population par sexe et âge ; et
- Obtenir d'autres données requises dans les questionnaires de la communauté.

Les tâches spécifiques pour l'étude de la Journée de l'Enfant comprennent les points suivants :

- Participation au développement du protocole d'étude,
- Développement des instruments de collecte de données,
- Formation pour les ONG locales, responsables de l'organisation des journées de l'enfance dans les paroisses, pour la conception expérimentale,
- Supervision des données collectées pendant la Journée de l'Enfant,
- Transcription des données,
- Saisie et gestion des données ; et
- Participation à l'analyse principale à réaliser à la fin de l'étude.

Tableau Annexe 2.1 Echantillons proposés pour l'évaluation de l'impact du projet de nutrition et de développement de la petite enfance de l'Ouganda

Catégorie	Utilisation de Vermifuge et éducation des parents		Non utilisation de Vermifuge et éducation des parents		Non utilisation de Vermifuge et non éducation des parents		Total
	Base	2 ^{ème} partie ^a	Base	2 ^{ème} partie ^a	Base	2 ^{ème} partie ^a	
Nombre de paroisses	25		25		25		
Période	Base	2 ^{ème} partie ^a	Base	2 ^{ème} partie ^a	Base	2 ^{ème} partie ^a	
Nombre de ménages	750	700	750	700	750	700	2.250
Nombre d'enfants pesés à la Journée de l'enfant ^b	5.000	5.000	5.000	5.000			20.000

(Le tableau continue sur la page suivante)

Catégorie	Utilisation de Vermifuge et éducation des parents		Non utilisation de Vermifuge et éducation des parents		Non utilisation de Vermifuge et non éducation des parents		
	1.500	1.395	1.500	1.395	1.500	1.395	
Nombre d'enfants âgés de 0 à 5,99 ans avec l'anthropométrie à la maison (deux par famille) ^c							11.580
Nombre d'enfants ayant reçu le test cognitif : test pour tous les enfants âgés de 4 à 5,99 ans dans les ménages	500 ^d	465 ^e	500	465	500	465	2.895
Nombre d'enfants âgés de 6 à 7,99 ans ayant reçu les tests cognitifs et d'anthropométrie	Sous-groupe ^f		Sous-groupe		Sous-groupe		Sous-groupe
Taux d'inscription scolaire	25	25	25	25	25	25	25
	Communitaires	Communitaires	Communitaires	Communitaires	Communitaires	Communitaires	Communautaires

Suppose une petite perte d'usure de 8 pour cent en deux ans.

- Suppose qu'environ 200 enfants assisteront à chaque Journée de l'enfant.
- Suppose qu'environ 200 enfants assisteront à chaque Journée de l'enfant
- Supposition de 2 enfants par famille, mais des familles seront recrutées si elles n'ont pas d'enfants en dessous de 6 ans. La famille se réfère ici à une paire qui se compose d'une mère (ou de la remplaçante) et de l'enfant.
- C'est un maximum; le nombre réel peut changer selon le nombre d'enfants de quatre à cinq ans pris en compte.
- Suppose la même perte de 8 pour cent sur 2 ans; seuls les enfants dont les parents ont été interviewés seront testés.
- Le nombre sera un sous-groupe d'enfants dans la tranche d'âge dont les parents ont été interviewés. Ils seront liés au résultat précédent. Quoique le nombre d'enfants testés ait augmenté au cours de la 2ème partie, la période consacrée aux interviews peut diminuer parce que beaucoup d'informations ne nécessiteront pas d'être évaluée à nouveau. Il est également possible que la taille du groupe soit réduite.

Source : Document de Projet de la Banque Mondiale

Etude de Validité. En plus de ce qui précède, une petite étude longitudinale sera ajoutée pour examiner la validité escomptée des mesures préscolaires, des résultats scolaires après la première année d'école. Dans l'enquête de base, deux enfants par communauté âgés de 6 à 6,9 ans (non encore scolarisés) seront testés, pour N = 150. Ces enfants seront situés en post-test et auront subi un test de résultat scolaire 2 ans plus tard, aux âges de 8 à 8,99 ans.

Programme des tâches. Le projet d'emploi du temps pour le programme des tâches est comme suit :

Mois 1. Commencer le processus de construction des indicateurs du développement cognitif conjointement avec le consultant international et en accord avec les termes de référence. Ce processus peut prendre 6 mois.

Mois 2. Pré Test initial et révision du questionnaire.

Mois 5. Commencer l'établissement de la liste des ménages pour l'échantillonnage. Cette étape dépend de la sélection des ONG principales et locales. Elle ne peut être entreprise sans que le Bureau de coordination du projet et les ONG aient sélectionné les paroisses où seront organisées les Journées de l'Enfant et aient donc sélectionné les sites pour abriter le programme initial d'éradication des vers. Dans le même temps, le questionnaire devra être traduit et le terrain testé à nouveau.

Mois 7. Commencer la collecte des données aux kermesses de l'enfant pour l'étude d'éradication des vers. Les données devront être rassemblées à ces kermesses à intervalles de six mois. Comme mentionné ci-dessus, la programmation de cette étape dépend de la sélection des ONG, principales et locales.

Mois 8. Formation du personnel sur le terrain pour l'enquête des ménages et démarrage de l'enquête. Cette dernière devrait prendre environ trois à quatre mois en fonction du nombre d'équipes recrutées. Les entrées de données devront correspondre à celles recueillies.

Mois 14. Analyse préliminaire des données de base. Ceci se déroulera d'une manière constante.

Mois 20. Le personnel du cabinet d'étude se réunit avec l'Université d'Oxford pour participer à l'analyse des données préliminaires.

Mois 20-36. Collecte des données du deuxième tour pour l'étude sur les vers. Les enquêtes des ménages à mi-parcours et finale se feront deux à quatre années après les données de base.

Soutien au Cabinet d'étude

Aucun soutien spécifique ne sera accordé au Cabinet d'étude pour effectuer ses missions. Il est recommandé aux cabinets d'étude d'inclure dans leurs propositions toutes les exigences en vue de la mise en œuvre effective de la mission.

Exemple II : Evaluation de l'impact des Pistes rurales : Niveau de référence de 1997 au Vietnam¹

Termes de Référence : Enquête de base sur l'étude de l'impact des Pistes rurales

I. Généralités

L'objectif de l'étude est d'évaluer l'impact du projet de transport rural financé par la Banque Mondiale sur le niveau de vie de la population au Vietnam. Ce projet qui a démarré en 1997, a été mis en œuvre dans 15 régions pauvres sur une période de trois à cinq années. L'étude sera axée globalement sur la manière dont les facteurs déterminants du niveau de vie changent dans le temps dans les communes qui reçoivent les interventions du projet routier en comparaison avec les communes qui n'en reçoivent pas. Pour ce faire, il est nécessaire de procéder à la collecte de données de base d'avant-projet à la fois pour les secteurs du projet ("traitement") et les secteurs de contrôle ("non traitement") et un nombre d'étapes ultérieures de collecte de données pour les données de post intervention, à intervalles de deux ans. Une base de données détaillée de la commune sera créée en partie par l'utilisation des archives rassemblées annuellement au niveau de la commune. A ces dernières seront ajoutées les données antérieures sur la commune et les diverses autres données supplémentaires principales collectées. Une brève enquête de la zone aidera à mettre les données de la commune dans le contexte. Finalement 10 à 15 ménages seront sélectionnés de façon aléatoire sur les listes de ménages de la commune et un petit questionnaire sur les ménages sera administré. L'étude sera entreprise dans 6 des 15 provinces qui bénéficieront du projet. Les 6 provinces représenteront les 6 régions géographiques du Vietnam. Un échantillonnage d'environ 200 des communes où le projet sera exécuté et de communes non concernées sera élaboré. Six équipes seront constituées pour examiner simultanément chaque province. L'enquête devrait démarrer en avril et se terminer vers le mois d'août. Les données devraient être disponibles en octobre ou novembre.

II. Conception de l'enquête

Echantillonnage. Il se fera à trois niveaux :

1. Les Provinces : Les 15 provinces de projet sont situées dans les six régions géographiques du Vietnam. Les critères de sélection des provinces d'étude seront les suivants : (a) une province sera sélectionnée dans chaque région géographique ; et (b) quand il y a plus d'une province de projet possible dans une région géographique, une sélection sera faite au hasard.

2. Les communes : Le but est d'examiner 200 communes ou plus qui sont sélectionnées au hasard. Environ la moitié ou moins (mais pas plus) devait être des communes ayant des projets de liaison routière, et le reste pour le contrôle. Une liste des communes non concernées par le projet sera élaborée dans les six provinces (ou alternativement une liste pour chaque province) et un échantillon sera tiré au hasard. De même une liste de toutes les communes bénéficiant de projets routiers sera élaborée dans les six provinces (ou province). Cela peut être plus d'une commune par liaison routière ; toutes les communes seront incluses dans le même cadre d'échantillonnage. Un échantillon sera également tiré au

¹ Ces termes de référence ont été élaborés par Dominique van de Walle.

IV. Divers

Equipement. Le matériel acheté pour les besoins du projet sera à la disposition du projet aussi longtemps que durera l'étude (à travers d'autres tournées) mais au cas où le matériel n'est pas utilisé par l'équipe, la direction du projet sera responsable de l'entreposer pour l'usage de l'équipe.

Déboursments du budget. Le budget alloué pour l'étude (à l'exclusion des paiements de l'enquêteur principal, qui recevra des traitements mensuels) sera déboursé en trois tranches. La première, représentant 20% du total du contrat sera payée dès la signature du contrat. Le deuxième acompte, d'un montant de 50% du budget total, ne sera déboursé que lorsque les questionnaires de la communauté, des ménages et des régions sont finalisés et approuvés par le Responsable du projet à la Banque Mondiale. La prévision de la fin de cette tâche se situe autour du mois de mars. Le troisième et dernier acompte sera déboursé à la fin juillet ou à mi-chemin de la collecte des données. Les détails de prévision budgétaire figurent dans le Tableau Annexe 2.II.2.

Tableau annexe 2.II.2. Prévisions budgétaires pour l'étude

	Nombre	Durée	Montant (US\$)	Total (US\$)
Enquêteur principal	1	9 mois	1.000	9.000
Expert d'enquête	6	6 mois	400	14.400
Frais de mission pour six enquêteurs, six guides locaux et interprètes	12	125 jours	8	12.000
Voiture et autres moyens de transport pour six équipes d'enquêteurs	6	125	40	30.000
Location de voiture pour l'enquêteur principal	1	30 jours	50	1.500
Billets d'avion Hanoi-Ho Chi Minh-Hanoi pour les 3 enquêteurs principaux (provinces du sud) (3 voyages aller-retour)	6		200	1.200
Formation des enquêteurs - Paiement - Déplacement sur le terrain - Frais de mission	12	1 semaine 3 jours/ 3 voitures 3 jours	50 50 8	1.338
Essai des questionnaires sur le terrain (communautés nord et sud)	1	2 semaines		2.000
Traitement de données et saisie	2	2 mois	200	800
Matériels d'enquête				2.000
Communication (fax, téléphone, e-mail, photocopie)				2.000
Equipement : • Ordinateur (PMU 18), • Imprimante (PMU 18), • Machine fax (équipe), • Ordinateur portable (équipe)	1 1 1 1		1.700 1.000 500 1.800	5.000

(Le tableau continue sur la page suivante)

Traduction (questionnaire, manuels, documentation) 200 pages	8/page	1.600
Tirage, photocopie		800
Imprévus		1.362
Total		85.000

Termes de référence : Superviseur de l'enquête ou enquêteur principal

I. Description des tâches

Le superviseur de l'enquête à l'intérieur du pays ou l'enquêteur principal sera responsable de l'enquête de base pour l'étude à l'intérieur du Vietnam. Le superviseur aura la responsabilité de rendre l'information disponible au niveau de la commune ; d'aider à la révision et à la finalisation des questionnaires au niveau de la zone, de la commune et des ménages ; de tester les questionnaires sur le terrain ; d'incorporer les corrections aux questionnaires ; de faire les arrangements en vue de faire traduire les questionnaires ; de recruter des assistants et les former ; de planifier la logistique des travaux sur le terrain ; de préparer la documentation de l'exécution des enquêtes et les questionnaires ; de superviser la réalisation des enquêtes et assurer le contrôle de la qualité du travail ; de superviser le nettoyage des données du projet et prendre des dispositions pour le traitement et la saisie des données. Cette personne assurera également la liaison avec la Division des projets au Ministère du transport (PMU 18), la Mission Résidente de Banque Mondiale, le représentant à Hanoi de l'Agence Canadienne pour le Développement International, et le responsable du projet à la Banque Mondiale à Washington. La personne dépendra directement de ce responsable du projet. Elle commencera son travail dès janvier 1997; le contrat peut être établi pour une période de neuf mois à un taux de salaire mensuel de \$1.000.

II. Tâches Spécifiques

Les tâches spécifiques comprennent les points suivants :

1. Assumer la responsabilité pour le recrutement, préparer les termes de référence détaillés, former et superviser les 6 assistants principaux qui travailleront avec les assistants locaux (si possible venant du bureau local des transports) sur le terrain et être responsable de la collecte des données au niveau de la région, de la commune et des ménages ;
2. Explorer la disponibilité des données au niveau de la commune et travailler en étroite collaboration avec le responsable de la Banque Mondiale pour élaborer les versions finales des questionnaires ;
3. Entreprendre un essai des questionnaires sur le terrain à la fois dans les communes rurales du sud et du nord; présenter un rapport sur les problèmes potentiels et les révisions nécessaires ; réviser les questionnaires en cas de besoin;
4. Prendre des dispositions pour que les questionnaires soient traduits, imprimés et photocopiés (les versions finales des questionnaires seront disponibles aussi bien en anglais qu'en vietnamien);
5. Sélectionner les six provinces devant être incluses dans l'enquête de sorte qu'il y ait une province par région géographique – quand il y a plus d'une province du type désiré, la sélection se fait au hasard; élaborer un échantillonnage au hasard d'environ 200 communes rurales dans les six provinces, en incluant celles avec des projets pour moitié environ et le reste avec celles non concernées par les projets ;
6. Planifier toute la logistique des travaux sur le terrain, y compris les dispositions pour le transport, les chauffeurs, les permis, le programme des enquêtes des communes et informer les autorités communales de l'arrivée des équipes et de leur objectif;
7. Participer à la réalisation des enquêtes, en supervisant alternativement les équipes; assurer un contrôle de qualité; identifier les problèmes affectant le déroulement des enquêtes, contrôler la qualité et s'assurer que les données rassemblées sont complètes, suggérer des moyens pour résoudre les problèmes et les mettre en pratique après consultation avec le responsable de l'étude ;

8. S'assurer que des enquêtes futures peuvent reproduire l'enquête de base, ce qui exige (a) la préparation de la documentation : détails de toute la conception de réalisation de l'enquête et de la logistique (le mode d'échantillonnage des provinces effectué, des communes et des ménages ; le mode de formation et d'organisation des équipes ; le mode d'organisation du travail sur le terrain ; le procédé adopté lorsqu'un site sélectionné se révélait inaccessible ou lorsqu'un ménage sélectionné n'a pas été retrouvé ; les problèmes, les questions soulevées et les solutions apportées) ; et (b) préparer un manuel détaillé sur les définitions des termes (par exemple, le chômage, le revenu, l'occupation principale, l'enfant ou l'adulte, la distance), les unités, les montants de devise, les codes utilisés dans les questionnaires ; la manière dont les questionnaire doivent être administrés et la population cible ; la manière dont les prix ont été relevés et autres ; la première étape devrait s'assurer que les suivantes peuvent reproduire l'organisation de base et les détails logistiques et la dernière devrait être utilisée dans la formation des enquêteurs et pour leur travail aussi bien que pour aider les futurs utilisateurs des données (des versions anglaises et vietnamiennes seront disponibles) ;
9. Fournir l'équipement nécessaire comme spécifié dans le budget ;
10. Etablir de bonnes relations et assurer une collaboration étroite avec la direction du projet du Ministère des Transports (PMU 18). Tenir le Ministère informé de l'étude et contrôler les développements du projet ; superviser la constitution d'une base de données à partir des données spécifiques du projet (le responsable de la Banque Mondiale identifiera les données à inclure) ;
11. Arranger et superviser le traitement et la saisie des données en utilisant le programme de saisie des données fourni ; et
12. Assurer la liaison et communiquer souvent avec le responsable du projet.

¹ Ces termes de référence ont été élaborés par Dominique van de Walle.

Annexe 3

Budget type d'une évaluation de l'impact d'un programme alimentaire scolaire

Phase I : Juillet 1999 – Décembre 2000 ^a

Proposition de recherche sur l'alimentation au niveau des écoles – Evaluation de base et d'échantillon
(Juillet 1999 – Décembre 2000)

Projet de budget – 14 / 7 / 1999 – US\$

	<i>Personnel semaine / Activité</i>		<i>Source de financement / coûts</i>			
	AF2000	AF2001	Petit Déjeuner	Logement. RPO	Autres	Total
Personnel de la Banque mondiale						
Economiste	4	2	17.640			
Expert en évaluation	5	3	23.520			
Expert en nutrition	5	3	23.520			
Pairs de critique	0.2	0.2	1.948			
Pairs de critique	0.2	0.2	1.948			
						68.577
Personnel FES						
Coordonnateur de Projet	4	4			12.000	
						12.000

(Le tableau continue sur la page suivante)

IV. Divers

Équipement. Le matériel acheté pour les besoins du projet sera à la disposition du projet aussi longtemps que durera l'étude (à travers d'autres tournées) mais au cas où le matériel n'est pas utilisé par l'équipe, la Direction du projet sera responsable de l'entreposer pour l'usage de l'équipe.

Déboursments du Budget. Le budget alloué pour l'étude (à l'exclusion des paiements de l'enquêteur principal, qui recevra des traitements mensuels) sera déboursé en trois tranches. La première, représentant 20% du total du contrat sera payée dès la signature du contrat. Le deuxième acompte, d'un montant de 50% du budget total, ne sera déboursé que lorsque les questionnaires de la communauté, des ménages et des régions sont finalisés et approuvés par le Responsable du projet à la Banque Mondiale. La prévision de la fin de cette tâche se situe autour du mois de mars. Le troisième et dernier acompte sera déboursé à la fin juillet ou à mi-chemin de la collecte des données. Les détails de prévision budgétaire figurent dans le Tableau Annexe 2.II.2.

Tableau annexe 2.II.2. Prévisions budgétaires pour l'étude

	Nombre	Durée	Montant (US\$)	Total (US\$)
Enquêteur principal	1	9 mois	1.000	9.000
Expert d'enquête	6	6 mois	400	14.400
Frais de mission pour six enquêteurs, six guides locaux et interprètes	12	125 jours	8	12.000
Voiture et autres moyens de transport pour six équipes d'enquêteurs	6	125	40	30.000
Location de voiture pour l'enquêteur principal	1	30 jours	50	1.500
Billets d'avion Hanoi-Ho Chi Minh-Hanoi pour les 3 enquêteurs principaux (provinces du sud) (3 voyages aller-retour)	6		200	1.200
Formation des enquêteurs	12			1.338
- Paiement		1 semaine	50	
- Déplacement sur le terrain		3 jours/ 3 voitures	50	
- Frais de mission		3 jours	8	
Essai des questionnaires sur le terrain (communautés nord et sud)	1	2 semaines		2.000
Traitement de données et saisie	2	2 mois	200	800
Matériels d'enquête				2.000
Communication (fax, téléphone, email, photocopie)				2.000
Équipement :				5.000
• Ordinateur (PMU 18),	1		1.700	
• Imprimante (PMU 18),	1		1.000	
• Machine fax (équipe),	1		500	
• Ordinateur portable (équipe)	1		1.800	

(Tableau continue sur la page suivante)

EXEMPLE DE TERMES DE REFERENCE		150
Traduction (questionnaire, manuels, documentation) 200 pages	8/page	1.600
Tirage, photocopie		800
Imprévus		1.362
Total		85.000

Termes de référence : Superviseur de l'enquête ou enquêteur principal

I. Description des tâches

Le superviseur de l'enquête à l'intérieur du pays ou l'enquêteur principal sera responsable de l'enquête de base pour l'étude à l'intérieur du Vietnam. Le superviseur aura la responsabilité de rendre l'information disponible au niveau de la commune ; d'aider à la révision et à la finalisation des questionnaires au niveau de la zone, de la commune et des ménages ; de tester les questionnaires sur le terrain ; d'incorporer les corrections aux questionnaires ; de prendre les dispositions en vue de faire traduire les questionnaires ; de recruter des assistants et les former ; de planifier la logistique des travaux sur le terrain ; de préparer la documentation de l'exécution des enquêtes et les questionnaires ; de superviser la réalisation des enquêtes et assurer le contrôle de la qualité du travail ; de superviser le nettoyage des données du projet et prendre des dispositions pour le traitement et la saisie des données. Cette personne assurera également la liaison avec la Division des projets au Ministère du transport (PMU 18), la Mission Résidente de Banque Mondiale, le représentant à Hanoi de l'Agence Canadienne pour le Développement International, et le responsable du projet à la Banque Mondiale à Washington. La personne dépendra directement de ce responsable du projet. Elle commencera son travail dès janvier 1997; le contrat peut être établi pour une période de neuf mois à un taux de salaire mensuel de \$1.000.

II. Tâches Spécifiques

Les tâches spécifiques incluent les points suivants :

1. Assumer la responsabilité pour le recrutement, préparer les termes de référence détaillés, former et superviser les 6 assistants principaux qui travailleront avec les assistants locaux (si possible venant du bureau local des transports) sur le terrain et être responsable de la collecte des données au niveau de la région, de la commune et des ménages ;
2. Explorer la disponibilité des données au niveau de la commune et travailler en étroite collaboration avec le responsable de la Banque Mondiale pour élaborer les versions finales des questionnaires ;
3. Entreprendre un essai des questionnaires sur le terrain à la fois dans les communes rurales du sud et du nord; présenter un rapport sur les problèmes potentiels et les révisions nécessaires ; réviser les questionnaires en cas de besoin;
4. Prendre des dispositions pour que les questionnaires soient traduits, imprimés et photocopiés (les versions finales des questionnaires seront disponibles aussi bien en anglais qu'en vietnamien);

5. Sélectionner les six provinces devant être incluses dans l'enquête de sorte qu'il y ait une province par région géographique – quand il y a plus d'une province répondant aux critères de sélection, la sélection se fait au hasard; élaborer un échantillonnage au hasard d'environ 200 communes rurales dans les six provinces, en incluant celles avec des projets pour moitié environ et le reste avec celles non concernées par les projets ;
6. Planifier toute la logistique des travaux sur le terrain, y compris les dispositions pour le transport, les chauffeurs, les per diems, le programme des enquêtes des communes et informer les autorités communales de l'arrivée des équipes et de leur objectif;
7. Participer à la réalisation des enquêtes, en supervisant alternativement les équipes; assurer un contrôle de qualité ; identifier les problèmes affectant le déroulement es enquêtes, contrôler la qualité et s'assurer que les données rassemblées sont complètes, suggérer des moyens pour résoudre les problèmes et les mettre en pratique après consultation avec le responsable de l'étude;
8. S'assurer que des enquêtes futures peuvent reproduire l'enquête de base, ce qui exige (a) la préparation de la documentation : détails de toute la conception de réalisation de l'enquête et de la logistique (le mode d'échantillonnage effectué des provinces, des communes et des ménages ; le mode de formation et d'organisation des équipes ; le mode d'organisation du travail sur le terrain ; le procédé adopté lorsqu'un site sélectionné s'est révélé inaccessible ou lorsque qu'un ménage sélectionné n'a pas été retrouvé; les problèmes, les questions soulevées et les solutions apportées) ; et (b) préparer un manuel détaillé sur les définitions des termes (par exemple, le chômage, le revenu, l'occupation principale, l'enfant ou l'adulte, la distance), les unités, les montants de devise, les codes utilisés dans les questionnaires ; la manière dont les questionnaire doivent être administrés et la population cible ; la manière dont les prix ont été relevés et autres; la première étape devrait s'assurer que les suivantes peuvent reproduire l'organisation de base et les détails logistiques et la dernière devrait être utilisée pour la formation des enquêteurs et pour leur travail aussi bien que pour aider les futurs utilisateurs des données (des versions anglaise et vietnamienne seront disponibles) ;
9. Fournir l'équipement nécessaire comme spécifié dans le budget ;
10. Etablir de bonnes relations et assurer une collaboration étroite avec la direction du projet du Ministère des Transports (PMU 18). Tenir le Ministère au courant de l'étude et contrôler les développements du projet; superviser la constitution d'une base de données à partir des données spécifiques du projet (le responsable de la Banque Mondiale identifiera les données à inclure);
11. Arranger et superviser le traitement et la saisie des données en utilisant le programme de saisie des données fourni ; et
12. Assurer la liaison et communiquer souvent avec le responsable du projet.

Consultants**Internationaux**

Evaluation de la situation				7.000	
(voyage compris)					
Développement du test cognitif (voyage compris)			6.000		
Spécialiste			2.000		
d'échantillonnage					
Etude de rentabilité			25.000		
					40.000

Cabinet régional de consultants ^b

Conception,			42.000		
échantillonnage, administration					
Travail sur le terrain			25.000		
Traitement de données			3.500		
Analyse			30.000		
					100.500

Voyage

Déplacements	4	2	12.000		
					12.000

Imprévus

Communication			1.000		
Logiciel			2.000		
Traduction			2.000		
					5.000

TOTAUX		68,577	150.150	19.000	238.077
---------------	--	--------	---------	--------	---------

Total requis de RAD : US\$ 150.500

Total requis du budget de la Banque: US\$68.577

Total fourni par des ressources extérieures : US\$19.000

- a. Les prévisions budgétaires de la phase II de l'évaluation ne sont pas incluses dans cette proposition.
- b. Le détail de ces coûts figure sur la page suivante.

Prévisions Budgétaires – Collecte de données locales et analyse pour la phase I

Evaluation de l'impact du programme d'alimentation en milieu scolaire

Coûts en US\$

	<i>Nombre de personnes</i>	<i>Nombre de personnes par semaine</i>	<i>Taux par semaine</i>	<i>Total</i>
Professionnels				
Directeur	1	12	2.000	24.000
Spécialiste de l'éducation	1	8	1.500	12.000
Spécialiste de la nutrition	1	8	1.500	12.000
Statisticien./ échantillonnage	1	12	750	9.000
Responsable du travail sur le terrain	1	8	750	6.000
Programmeur	1	10	300	3.000
Superviseur de traitement	1	8	300	2.400
de données				
Assistant – enquêtes	1	10	100	1.000
Assistant – anthropométrie	1	10	100	1.000
Assistant – test cognitif	1	10	100	1.000
Contrôle de qualité des données	1	8	100	800
Sous-total – Personnel cadre				72.200
Personnel de terrain				
Superviseur	4	6	200	4.800
Chargé du test cognitif	4	6	120	2.880
Anthropométriste	4	6	120	2.880
Intervieweur	4	6	120	2.880
Chauffeur	4	5	100	2.000
Equipement de terrain				
	Personnes/unités		Coût par semaine ou unité	
Véhicules (4 véhicules	4	5	350	7.000
Pour 5 semaines)				
Carburant (4 véhicules pour 5 semaines)	4	5	80	1.600
Balances, régisseurs (5 ensembles)	5		20	100
Matériel de test cognitif (pour 4 chargés de test)	4		20	80
Matériel d'enquête (pour 4 intervieweurs)	4		20	80
Sous-total travail de terrain				24.300
Traitement de données				
	Personnes			
Codage de données	3	7	75	1.575
Saisie des données	4	7	75	2.100
Sous-total traitement de données				3.675
TOTAL				100.175

Annexe 4

Indicateurs d'impact – Evaluation du fonds social d'investissement en Bolivie

Développé en nombre 1997

I. Enseignement formel– Ecoles de type "A" et de type "B"

(niveaux multiples et réguliers)

1. Indicateurs d'impact final

Réussite dans les tests de Mathématique et langues (a)
Taux de redoublement
Taux d'abandon
Inscription
Niveau d'instruction
Demande pour l'éducation (pourcentage d'élèves exclus de l'école) ^a

2. Indicateurs d'impact intermédiaire

Régularité des élèves
Régularité des enseignants
Temps alloué pour les élèves/ temps passé à étudier
Méthode d'enseignement en classes
Mouvement du personnel enseignant

3. Indicateurs d'intervention

Infrastructure
Rapport étudiants/ classe
Nombre de classes en "bon état"
Nombre de classes manquantes
Disponibilité de zone multi-fonctionnelle
Disponibilité de services de base

- Electricité
- Alimentation en eau
- Type d'installations sanitaires ; condition des services sanitaires

Meubles

- Nombre d'étudiants / table
 - Nombre de bureaux d'enseignant / classe
 - Nombre de chaises d'enseignant / classe
 - Nombre de "tableaux noirs adéquats" / classe
 - Nombre d'étagères / classe
 - Nombre de textes et matériel didactique
 - Nombre de manuels / élève
 - Qualité des textes de mathématiques, de langue, de sciences humaines et de sciences naturelles
 - Disponibilité de textes d'enseignants
 - Disponibilité et état des cartes et plans
 - Jeux didactiques par cycle scolaire (pré-base, base, intermédiaire)
 - Disponibilité d'un boulier
- Indicateurs de réforme de l'éducation ^b

4. **Facteurs affectant les résultats non liés au Projet SIF (exogènes)**

- Nutrition
- Disponibilité de programme de petit déjeuner à l'école
- Frais de scolarité
- Caractéristique des enseignants
- Profil de Formation
- Années de service
- Formation reçue
- Méthodes d'enseignement utilisées (en classe)
- Formation reçue, par thème et cours
- Pratiques d'évaluation des élèves (fréquence de devoirs de maison et correction)
- Evaluation des enseignants par les élèves
- Raisons d'abandon
- Elèves exclus de l'école
- Distance entre le domicile et l'école
- Ratio élèves/ enseignant

5. **Indicateurs d'identification**

- La priorité est-elle donnée à l'école par la Réforme de l'Education ?
- Frais encourus par composante du projet
- Dépenses réelles par composante du projet

II - Santé

1. Indicateurs d'impact final ^c
 - Taux de mortalité des tout petits
 - Taux de mortalité infantile
 - Taux d'incidence et de prévalence des principales maladies
 - Prédominance de la malnutrition (général, léger, modéré et grave)
2. Indicateurs d'impact intermédiaire
 - Utilisation de centres de santé gouvernementaux
 - Existence du vaccin antitétanique

- Lieu où le vaccin a été fait
- Prévalence du contrôle prénatal
 - Nombre de contrôles prénatals effectués
 - Qualité du contrôle
- Prévalence des accouchements effectués dans les centres de santé
 - Qualité des soins
- Prévalence des accouchements à domicile effectués par le personnel médical.
- Taille à la naissance
- Poids à la naissance
- Mesures anthropométriques
 - Lieu où la mesure a été prise
 - Age au moment de la première mesure
- Fréquence de maladie et prévalence d'immunisation par nombre de doses reçues
 - Polio
 - Diphtérie -Tétanos-Coqueluche (DPT)
 - Rougeole
 - Vaccin anti-tuberculose
- Connaissance des endroits où se rendre pour l'immunisation
- Fréquence d'accès de toux et traitement
- Fréquence d'accès de diarrhée et traitement
- Prévalence de la connaissance de l'utilisation des paquets de réhydratation
- Connaissance des cliniques à prévalence de grossesse
- Taux de grossesses à haut-risque
- Prévalence des habitudes de bonne hygiène et d'utilisation de l'eau
- Durée de l'allaitement

3. Indicateurs d'intervention ^d

- Qualité de l'infrastructure par type de centre de santé
- Disponibilité de services de base dans les centres de santé (eau courante, réseau d'égout et électricité)
- Adéquation de l'infrastructure selon les normes par type de centre de santé
- Adéquation de l'équipement selon les normes par type de centre de santé
 - Nombre de lits dans le centre de santé
- Disponibilité des médicaments essentiels par type de centre de santé
- Disponibilité d'instruments médicaux essentiels par type de centre de santé
- Disponibilité de fournitures essentielles par type de centre de santé

4. Facteurs affectant les résultats non liés au Projet SIF (exogène)

- Caractéristiques du ménage
 - Qualité du lieu d'habitation
 - Type d'habitation
- Services de base du lieu d'habitation
 - Electricité
 - Eau courante
 - Type d'installations sanitaires
- Accessibilité aux services de base
 - Distance séparant le lieu d'habitation et le centre de santé le plus proche
 - Distance séparant les installations sanitaires et la source
 - Distance séparant le lieu d'habitation et le point principal d'alimentation en eau
 - Heures de disponibilité d'eau par jour

Suffisance de la quantité d'eau par jour
 Disponibilité de la quantité d'eau tout au long de l'année
 Coût de la consultation dans un centre de santé
 Perception du chef de famille sur la qualité :
 du "service" dans le centre de santé fréquenté par sa famille,
 de "l'infrastructure" du centre de santé fréquenté par sa famille,
 de la "disponibilité des médicaments" dans le centre de santé fréquenté par sa famille.

Dépenses de ménage

Caractéristiques personnelles des membres de la famille

Age

Langue

Niveau d'instruction

Occupation

Caractéristiques géographiques

District de santé

Zone de santé

Secteur de santé

Province

Localité

Ressources humaines existantes dans le centre de santé (médecins, odontologistes, nutritionnistes, infirmières, aide-infirmières, techniciens, personnel administratif)

Population sous la zone de couverture du centre de santé par groupes d'âge

Frais de consultation dans le centre de santé

Interventions de santé non financées par le SIF

5. Indicateurs d'identité

Frais encourus par composante de projet

Dépenses réelles par composante de projet

III. Eau

1. Indicateurs d'impact final ^e

Taux de mortalité des tout petits

Taux de mortalité infantile

Taux de fréquence et de prévalence de diarrhée dans la famille

Prévalence de la malnutrition (générale, légère, modérée et grave)

2. Indicateurs d'impact intermédiaire

Fréquence et traitement de diarrhées dans les centres de santé

Prévalence et connaissance de l'utilisation des paquets de réhydratation orale

Prévalence des habitudes de bonne hygiène et d'utilisation d'eau

3. Indicateurs d'intervention (des ressources)

Prévalence de la formation en matière de santé

Accessibilité aux services de santé

Principale source d'eau

Existence et type d'installations sanitaires

Distance séparant les installations sanitaires et la source
 Distance séparant le lieu d'habitation et la principale source d'eau
 Heures de disponibilité d'eau par jour
 Suffisance de quantité d'eau par jour
 Disponibilité d'eau pendant l'année
 Quantité d'eau utilisée par la famille ^a
 Qualité de l'eau ^a

4. Facteurs affectant les résultats non liés au Projet SIF (exogènes)

Utilisation des centres de santé du gouvernement (MSSP)
 Taille à la naissance
 Poids à la naissance
 Durée de l'allaitement
 Caractéristiques du ménage
 Qualité du lieu d'habitation
 Type d'habitation
 Accessibilité aux services de base
 Distance séparant la famille et le centre de santé le plus proche
 Frais de consultation dans le centre de santé
 Dépenses du ménage
 Caractéristiques personnelles des membres de la famille
 Age
 Langue
 Niveau
 Occupation

5. Indicateurs d'identification

Frais encourus par composante du projet
 Dépenses réelles par composante du projet

-
- a. Non considérée au départ.
 - b. A développer en coordination avec le personnel de la Réforme de l'Education ; seront considérées comme exogènes aux interventions à moins que les interventions SIF de la réforme d'éducation soient associées.
 - c. Taux général de morbidité, taux de natalité, taux global de fécondité, mortalité adulte et espérance de vie à la naissance supprimée.
 - d. Formation en matière de santé supprimée.
 - e. Taux général de mortalité, taux de naissance, taux global de fécondité, mortalité adulte (homme et femme), espérance de vie à la naissance, prévalence d'infections respiratoires aiguës et traitement d'accès de toux supprimé.

Annexe 5

Cadre Logique des récapitulatifs de conception du projet pour le document d'achèvement du projet ou le document d'évaluation du projet

<i>1. Hiérarchie des objectifs</i>	<i>Indicateurs clés de performance</i>	<i>Contrôle et évaluation</i>	<i>Hypothèse critique</i>
<p>Objectif de stratégie d'assistance au pays, liée à un secteur.</p> <p>Fournir une déclaration en une phrase sur l'objectif stratégique à long terme (tel que reflété dans le CAS) auquel le projet est destiné à contribuer.</p> <p>La déclaration doit décrire le développement significatif opéré dans</p> <p>le (s) secteur (s) concerné (s).</p>	<p>Indicateurs de Secteur :</p> <p>1. Les indicateurs accompagnant le secteur concerné par l'objectif CAS incluent des dispositions non financées par le projet mais qui peuvent être financées par la Banque comme partie d'autres travaux.</p> <p>2. En principe, l'emprunteur contrôlera ces indicateurs comme partie de bonne pratique de gestion.</p>	<p>Rapport du secteur et du pays :</p> <p>1. Cette colonne identifie le lieu où la vérification de chaque indicateur peut se trouver et le mode d'opération.</p> <p>2. Les indicateurs accompagnant le secteur concerné par l'objectif CAS sont généralement contrôlés et évalués grâce à de nombreux rapports du secteur ou du pays qui sont produits en dehors du programme.</p>	<p>(De l'objectif à la mission de la Banque)</p> <ul style="list-style-type: none"> • En supposant que l'objectif du CAS (indiqué dans la colonne à l'extrême gauche) soit atteint à long terme, lister d'autres suppositions complémentaires nécessaires pour lier cet objectif à la mission de la Banque (c'est-à-dire la réduction de la pauvreté). • Ces suppositions incluent souvent les conditions, actions ou réponses à l'extérieur du projet et en dehors du secteur.

(L'annexe 5 continue sur la page suivante)

Annexe 5 (suite)

Objectif de développement du projet:

1. Enoncer une déclaration en une phrase sur le changement de comportement attendu de la part du groupe ou des institutions bénéficiaires du programme à la fin de son exécution.

La réalisation de l'objectif sert comme simple test de demande pour les résultats du projet.

L'objectif doit être l'expression explicite d'un développement réaliste, spécifique, quantifiable et à la demande. Comme manuel de référence pour devenir l'objectif du projet, consulter le guide "Ce qu'il faut faire ou ne pas faire dans la définition d'un projet de développement" (appeler x 7065 ou envoyer un e-mail à l'adresse M&Ehelp@worldbank.org pour obtenir un exemplaire.)

Réalisation / Indicateurs

d'impact :

1. Les indicateurs se rapportent aux résultats escomptés, alors que l'impact ne sera visible que dans 5 ans ou plus après la fin de l'exécution du projet.

2. Les indicateurs d'achèvement (PDO) ne sont pas une réaffirmation de ceux au niveau des résultats.

3. La collecte de données pour quantifier les indicateurs est généralement financée par le projet.

Rapports de projet :

1. Ce chapitre identifie le lieu où les informations permettant la vérification de chaque indicateur peuvent être trouvées et le mode d'opération.

2. Les indicateurs accompagnant l'objectif de développement du projet sont généralement contrôlés et évalués à travers de nombreux rapports de projet, de rapports de mission de supervision et d'évaluation (à mi-parcours ou à la fin).

3. Lorsqu'une collecte de données est nécessaire, une mention spécifique doit être faite des méthodes et des responsabilités, qui peuvent inclure une demande de renseignements venant des bénéficiaires.

(De l'objectif de développement du projet à l'objectif CAS du secteur)

- En supposant que l'objectif de développement du projet soit atteint, faire la liste d'autres hypothèses nécessaires pour justifier la contribution du projet à l'objectif déclaré.
- Ces hypothèses se rapportent aux contributions supplémentaires d'autres projets, apports ou réponses des groupes ou institutions bénéficiaires qui sont pertinentes à la réalisation de l'objectif déclaré.

(L'annexe 5 continue à la page suivante)

Annexe 5 (suite)

Résultat de chaque composante :	Indicateurs de résultat :	Rapports de projet :	(Des résultats à l'objectif de développement du projet)
<p>1. Stipuler ici (au passé) la valeur ajoutée par la réalisation de chaque composante.</p> <p>2. Une déclaration correcte de la valeur ajoutée sera facile à mesurer (comme reflété dans les indicateurs de droite).</p> <p>3. Pour une simplicité et une clarté dans la logique, il doit y avoir une déclaration de résultat pour chaque composante correspondante.</p> <p>4. Chaque résultat doit correspondre en nombre à sa composante respective.</p> <p>5. L'équipe chargée du projet a généralement la responsabilité de s'assurer de la production de ces résultats comme partie</p>	<p>1. Les indicateurs de résultat ont des attributs de quantité, qualité et durée. Si la durée n'est pas stipulée, la fin du projet est supposée l'être.</p> <p>2. Les indicateurs de résultat incluent des mesures de rentabilité.</p> <p>3. La collecte des données pour mesurer les indicateurs de résultat est financée par le projet.</p> <p>4. Pour les projets complexes, un tableau séparé (un addendum à l'annexe 1) peut être utilisé pour fournir une liste plus détaillée des indicateurs de résultat.</p> <p>5. Il est préférable de n'avoir que peu d'indicateurs de résultat significatifs et faciles à mesurer que d'avoir un</p>	<p>1. Les indicateurs de résultat sont généralement contrôlés et évalués à travers plusieurs rapports de projet, rapports de missions de supervision et d'évaluation (à mi-parcours ou à la fin).</p> <p>2. Les sources de donnée pour le contrôle et l'évaluation de ces indicateurs incluent normalement les systèmes d'archivage administratifs et de gestion, et les rapports établis par le projet.</p>	<ul style="list-style-type: none">• En supposant que les données énoncées dans le cadre à l'extrême gauche soient réalisées à la fin du projet, faire la liste d'autre hypothèses complémentaires nécessaires pour réaliser les objectifs du projet.• Ces hypothèses peuvent englober des conditions, des changements de politique ou les réactions attendues de la part des groupes ou institutions bénéficiaires qui sont nécessaires au succès du projet.• Ces hypothèses sont critiques à la réalisation des

intégrante d'une bonne conception de projet et de sa bonne exécution, planification et livraison.

grand nombre d'indicateurs pour lesquels la collecte des données est problématique.

6. Les indicateurs de résultat font l'objet d'un commun accord avec l'emprunteur à l'étape PCD (quant à la disponibilité des données et la facilité de la collecte) et une base est obtenue avant l'évaluation.

objectifs énoncés du projet mais sont en dehors du contrôle direct du projet.

(L'annexe 5 continue sur la page suivante)

Annexe 5 (suite)

Les composantes/sous-composantes du projet :

1. Une composante est un ensemble de sous-composantes ou d'activités conçues pour constituer un seul projet.

2. Faire la liste de chaque composante de projet comme rubrique principale suivie des sous-composantes principales, s'il y en a, comme partie du projet.

Ressources du projet (budget pour chaque composante) :

1. Faire la liste des composantes en terme de coût total de chaque composante, y compris les imprévus (par exemple, US\$_____).

2. Pour des projets, grands ou complexes, les coûts des sous-composantes peuvent aussi être indiqués (en alinéa, pour les séparer des coûts pour les composantes).

Rapports de projet :

1. Les ressources sont généralement contrôlées à travers les rapports d'avancement et de déboursement (tous deux trimestriellement).

2. Les ressources sont généralement évaluées à travers les rapports de missions de supervision (deux par an) et les rapports d'audit (annuels).

(Des composantes du projet aux résultats du projet)

- Avec l'hypothèse que les composantes et les activités énoncées dans le cadre à l'extrême gauche soient réalisées avec succès, faire la liste d'hypothèses complémentaires nécessaires pour réaliser les rendements déclarés.
- Ces hypothèses sont des conditions en dehors du contrôle direct du projet et sont nécessaires si les résultats stipulés doivent être réalisés par le

- projet.
- Le projet en lui-même ne devrait pas dépenser de l'argent pour réaliser n'importe laquelle de ces conditions (puisque de telles hypothèses sont incluses dans les composantes elles-mêmes)

Source : Département des services opérationnels centraux, Banque Mondiale. Pour des exemples complets de cette annexe, visiter le site M&E Help du service Internet Web de la Banque à l'adresse <http://L.nts012/helhelpdesk.nsf>.

Annexe 6

Matrice d'Analyse : Evaluation de l'impact du Fonds social d'investissement d'urgence du Nicaragua – 1998

A. Cibler la pauvreté

164	EVALUATION DE L'IMPACT DES PROJETS DE DEVELOPPEMENT SUR LA PAUVRETE			
communautés et régions SF	<p>la communauté ou la région en dessous du seuil de pauvreté ou du niveau de consommation du très pauvre.</p> <ul style="list-style-type: none"> • Niveau de consommation moyen dans les communautés participant au fonds social comparé au niveau de consommation du pays • Indice de pauvreté (tel qu'utilisé par SF) 	<p>revenu et la consommation des ménages et l'identification des activités SF par communauté et par région</p> <p>Nécessite une enquête sur le revenu et la consommation des ménages et l'identification des activités du SF par communauté et par région</p> <p>Les cartes utilisent habituellement des mesures de procuration comme indice de pauvreté basé sur des variables mixtes.</p>	<p>faut des définitions semblables des seuils de pauvreté</p> <p>Les inconvénients sont qu'elles choisissent arbitrairement les indicateurs et les pondérations, et chaque pays a un indice différent. L'avantage est qu'elles fournissent souvent plus de</p>	<p>dans les communautés SF- pour uniquement l'éducation, la santé, l'eau et les projets d'installations sanitaires</p> <p>Sondage national des ménages (LSMS) dans les communautés- uniquement pour l'éducation, la santé, l'eau et les projets d'installations sanitaires</p> <p>Utilisation SF de la carte de la pauvreté à partir des données LSMS93 utilisant l'indice de pauvreté composite; mettra à jour l'utilisation de LSMS98 et les données de recensement pour prévoir la</p>
			données d'ensemble que les enquêtes sur les revenus et la consommation –	consommation à des niveaux sub-nationaux .

L'annexe 6 continue sur la page suivante

			<p>les deux peuvent être liées pour dériver les niveaux de consommation prévus à des niveaux désagrégés.</p> <p>Peut varier largement par type de projet SF</p>	
Seuils de pauvreté des bénéficiaires SF (ménages)	<ul style="list-style-type: none"> • Pourcentage des bénéficiaires en dessous du seuil de pauvreté ou dans une pauvreté extrême 	Enquête sur le revenu et la consommation qui sélectionne les bénéficiaires soit par taille de SF soit par sondage dans les communautés SF		Sondage national des ménages (LSMS) dans les communautés SF
	<ul style="list-style-type: none"> • Consommation moyenne des ménages bénéficiaires comparée à la moyenne nationale pour des ménages semblables par type de projet (par exemple, avec enfants à l'école primaire, accès à l'eau courante et utilisation de latrines) 	Enquête sur le revenu et la consommation qui sélectionne les bénéficiaires SF soit par taille soit par sondage dans les communautés SF; peut aussi faire un sondage dans les communautés correspondantes sans interventions SF		Sondage national des ménages (LSMS) dans les communautés SF

(L'annexe 6 continue à la page suivante)

Matrice (suite)

<i>Questions / Objets</i>	<i>Indicateurs généraux</i>	<i>Méthodologies</i>	<i>Commentaires</i>	<i>Source de données</i>
Distribution de ressources SF	<ul style="list-style-type: none"> • Pourcentage de projets de fond social et de ressources dans les groupes de bas niveau des régions 	Nécessite une méthodologie de classement consistante dans tous les pays		Nécessite une révision du système de classement et un recalibrage
Caractéristiques de politique institutionnelle qui affectent l'application du ciblage SF	<ul style="list-style-type: none"> • Utilisation d'une carte d'indice de pauvreté • Effort de promotion • Accès direct par les groupes bénéficiaires • Utilisation en commun des projets par les agences qui en font la demande • Décentralisation des bureaux • Ciblage de la dotation des ressources • Liste de sous-projets 	Développe des variables institutionnelles standards qui peuvent être utilisées pour expliquer les résultats visés, variables obtenues facilement du Fonds social (SF)		Information disponible au FS
Autres facteurs affectant l'application du ciblage	<ul style="list-style-type: none"> • Age du Fonds Social • "Capital social" de la communauté • Distance au siège du Fonds social • Niveau d'instruction le plus élevé des bénéficiaires • Présence d'intervention du gouvernement ou des ONG • Degré d'inégalité des revenus dans le pays 	Nécessite également des définitions standards des variables - variables obtenues du FS, des enquêtes sur des ménages (avec identification des bénéficiaires du FS) et des enquêtes nationales		Le seul indicateur peu sûr est "le capital social de la communauté "

(La matrice continue à la page suivante)

Comparaison d'alternatives	<ul style="list-style-type: none"> ● Pourcentage de projets FS et de ressources dans les basses couches des régions, comparé à d'autres programmes semblables et mécanismes de livraison 	Compare l'application du ciblage basé sur la localisation géographique ou les degrés de pauvreté des bénéficiaires, compte tenu de la conception de l'enquête, l'échelle du FS et autres programmes	Difficulté pour trouver des éléments de comparaison fiables, nécessité de rassembler une information séparée à partir des programmes de comparaison	Planifié pour une analyse de rentabilité
----------------------------	---	---	---	--

B. Bénéfices

<i>Questions/Sujets</i>	<i>Indicateurs généraux</i>	<i>Source de données/méthodologies</i>	<i>Commentaires</i>	<i>Etude de cas : Nicaragua</i>
Capital physique	<ul style="list-style-type: none"> ● Mesure dans laquelle les micro-projets répondent aux priorités des communautés 	Les enquêtes de communauté, l'évaluation des bénéficiaires ou l'enquête sur des ménages avec des sondages dans les zones FS		Evaluation des bénéficiaires couverts par le financement BIRD et enquête sur les facilités.
	<ul style="list-style-type: none"> ● La perception du niveau de bénéfice par les bénéficiaires et les améliorations du bien-être 	Enquête sur des ménages ou évaluation des bénéficiaires dans les communautés SF		Evaluation des bénéficiaires couverts par l'enquête sur les ménages et le financement BIRD
	<ul style="list-style-type: none"> ● Amélioration dans l'accès aux infrastructures sociales et économiques (avant et après) 	Enquête sur des ménages des bénéficiaires du FS	Nécessité d'avoir soit des questions de base soit des questions de rappel	Questions de rappel de l'enquête sur les ménages, possibles aussi ex ante à partir des études LSMS précédentes; peut comparer les bénéficiaires du SF avec les caractéristiques de la population nationale et les communautés correspondantes

(La matrice continue à la page suivante)

	<ul style="list-style-type: none"> ● Amélioration dans l'accès aux infrastructures sociales et économiques en parallèle avec des projets de comparaison 	Enquête sur des ménages du SF et des bénéficiaires de projet de comparaison	Nécessite d'avoir soit des questions de base, soit des questions de rappel; besoin de développer des indicateurs séparés par type de projet SF	Peut comparer les bénéficiaires SF avec les caractéristiques générales de la population nationale aussi bien que les communautés correspondantes
	<ul style="list-style-type: none"> ● Amélioration dans la qualité des infrastructures et des services (avant et après) 	Enquête sur les facilités et enquête sur des ménages, évaluation de la couverture par des bénéficiaires	Nécessite d'avoir soit des questions de base, soit des questions de rappel; besoin de développer des indicateurs séparés par type de projet SF	Pour l'éducation, la Santé, l'eau et les infrastructures sanitaires, rappel plus information historique des enquêtes de facilités et de l'évaluation SF ex ante.
	<ul style="list-style-type: none"> ● Amélioration dans la qualité des infrastructures et services en parallèle avec des projets de comparaison 	Enquête sur les facilités, enquête sur des ménages, évaluation de la couverture par des bénéficiaires (Fonds Social et projets de comparaison)	Nécessite d'avoir soit des questions de base, soit des questions de rappel; besoin de développer des indicateurs séparés par type de projet SF	Pour l'éducation la santé, l'eau et les infrastructures sanitaires, les facilités SF et non SF à travers les enquêtes de ménage et les enquêtes de facilités.
Capital humain	<ul style="list-style-type: none"> ● Amélioration de l'enseignement : fréquentation scolaire, années d'études, taux d'abandon et de redoublement (avant et après, et en parallèle avec les groupes de comparaison) ● Amélioration de la santé : par exemple, fréquence de maladie, mortalité infantile, malnutrition, augmentation de l'allaitement maternel (avant et après et en parallèle avec les groupes de comparaison) 	<p>Enquête sur des ménages et informations en provenance de l'école</p> <p>Enquête sur des ménages avec un module de santé, des mesures anthropométriques si la malnutrition est incluse.</p>		<p>Enquête sur des ménages et information en provenance de l'école pour les écoles et ménages SF et non SF</p>

(La matrice continue à la page suivante)

	<ul style="list-style-type: none"> • Amélioration de l'économie : augmentation de revenus, réduction de temps passé à chercher de l'eau, coût faible des services, augmentation de l'emploi (avant et après et en parallèle avec les groupes de comparaison) 	Enquête sur des ménages		Enquête sur des ménage pour les bénéficiaires SF et les communautés correspondantes non bénéficiaires
Capital social	<ul style="list-style-type: none"> • Augmentation de la capacité communautaire à faire face aux problèmes (en parallèle avec les groupes de comparaison) 	Enquête sur des ménages, enquête de communauté et/ou évaluation de bénéficiaires		Non traité
	<ul style="list-style-type: none"> • Augmentation du taux de participation dans les changements initiés par la communauté (en parallèle avec les groupes de comparaison) 	Enquête sur des ménages, enquête de communauté ou évaluation de bénéficiaires	Nécessite de développer des indicateurs	Information dans l'enquête de ménage sur la participation
C. Pérennité des bénéfices				
<i>Questions/Objets</i>	<i>Indicateurs généraux</i>	<i>Sources de données</i> <i>Méthodologies</i>	<i>Commentaires</i>	<i>Etude de cas : Nicaragua</i>
Pérennité des opérations	<ul style="list-style-type: none"> • Conditions dans lesquelles les projets SF fonctionnent après les interventions SF (durabilité absolue) 	Enquête de facilités	Peut avoir des informations complémentaires à partir de l'évaluation des bénéficiaires	Pour les enquêtes de projet d'éducation et de santé, avoir à la fois des bénéficiaires SF et des non SF.
	<ul style="list-style-type: none"> • Condition dans lesquelles les projets SF fonctionnent après les interventions SF face à des projets de comparaison (durabilité relative) 	Enquête de facilités	Peut tirer des informations complémentaires de l'évaluation des bénéficiaires	Pour les enquêtes de projet d'éducation et de santé, avoir à la fois des bénéficiaires SF et des non SF.

(La matrice continue à la page suivante)

Durabilité de la maintenance	<ul style="list-style-type: none"> ● Maintenance des infrastructures et des services (absolue) dans le temps 	Enquête de facilités	Possibilité d'avoir des informations supplémentaires de l'évaluation des bénéficiaires	Pour les enquêtes de projet de santé et d'éducation, avoir à la fois des bénéficiaires SF et des non SF
	<ul style="list-style-type: none"> ● Maintenance des infrastructures et des services dans le temps en parallèle avec des projets de comparaison (relative) 	Enquête de facilités	Possibilité d'avoir des informations supplémentaires à partir de l'évaluation des bénéficiaires	Pour les enquêtes de projet de santé et d'éducation, avoir à la fois des bénéficiaires SF et des non SF
	<ul style="list-style-type: none"> ● Qualité et quantité des infrastructures et services dans le temps 	Enquête de facilités et des ménages	Possibilité d'avoir des informations supplémentaires à partir de l'évaluation des bénéficiaires	Pour les enquêtes de projet de santé et d'éducation, avoir à la fois des bénéficiaires SF et des non SF
Durabilité de l'impact	<ul style="list-style-type: none"> ● Tendance des communautés SF à soumettre d'autres propositions (à SF et autres) dans le temps 	Base de données SF, enquête de communauté ou évaluation des bénéficiaires		Nécessiterait un travail complémentaire
Durabilité des effets communautaires	<ul style="list-style-type: none"> ● Participation de la communauté aux besoins d'infrastructures sociales et économiques dans le temps 	Enquête de communauté, enquête sur des ménages ou évaluation des bénéficiaires		Inclure dans la prochaine évaluation des bénéficiaires; évaluation de l'impact
D. Questions de rentabilité				
<i>Questions/Sujets</i>	<i>Indicateurs généraux</i>	<i>Source de données</i> <i>Méthodologies</i>	<i>Commentaires</i>	<i>Etude de cas : Nicaragua</i>
Rentabilité des micro-projets	<ul style="list-style-type: none"> ● Coût moyen par école nouvelle, poste de santé, système d'adduction d'eau avec des approches alternatives en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance, des ministères et des municipalités	Les changements de coût dans le temps et les projets de comparaison doivent être identiques	Données SF et non SF à partir de l'enquête des facilités. Les estimations de coûts non SF peuvent ne pas être fiables

(La matrice continue à la page suivante)

	<ul style="list-style-type: none"> • Coûts par unité : coût par mètre carré de construction, par kilomètre de route, et ainsi de suite en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance des ministères et municipalités	Peut calculer les moyennes SF. A inclure dans l'analyse de rentabilité
	<ul style="list-style-type: none"> • Coût moyen par bénéficiaire par projet SF type en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance des ministères et municipalités	Peut calculer les moyennes SF. A inclure dans l'analyse de rentabilité
Rentabilité des mécanismes de production	<ul style="list-style-type: none"> • Coût moyen d'emploi généré en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance des ministères et municipalités	Peut calculer les moyennes SF. A inclure dans l'analyse de rentabilité
	<ul style="list-style-type: none"> • Coût institutionnel SF (investissement et fonctionnement) comme part des projets SF en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance des ministères et municipalités	Nécessité de développer des définitions standards des coûts institutionnels; spécifier la période
	<ul style="list-style-type: none"> • Temps d'achèvement moyen en parallèle avec des projets de comparaison 	Base de données SF et informations en provenance des ministères et municipalités	Peut calculer les moyennes SF. A inclure dans l'analyse de rentabilité