

L'ÉVALUATION EN SITUATION REELLE < EN BREF >

Appliquer les approches d'Évaluation en situation réelle à tous les stades du processus d'évaluation

Extrait et traduction de la livre :

***RealWorld Evaluation :
Working under Budget, Time, Data and
Political Constraints,***

Michael Bamberger, Jim Rugh et Linda Mabry
Sage Publications, 2006

Traduction par Marine Armstrong

TABLE DES MATIERES

Introduction à cette version détachable	1
Cerner le champ de l'évaluation	1
Choisir le meilleur type d'évaluation parmi les choix possibles.....	9
Déterminer quelles méthodologies conviennent.....	17
Moyens pour renforcer les modèles d'évaluation en situation réelle	17
Doter l'évaluation en personnel de façon économique	23
Collecter les données efficacement	26
Analyser les données efficacement	29
Rapporter les constatations avec efficience et efficacité	32
Aider les clients à faire bon usage des constatations.....	35
Annexe : Principales figures et tableaux extraits des autres chapitres	36

Introduction à cette version détachable

Dans le Chapitre 16 du livre, nous examinons les façons dont les approches d'**Évaluation en situation réelle (ESR)** peuvent être appliquées à tous les stades de la conception et de la mise en œuvre d'une évaluation typique. Nous indiquons les problèmes qui peuvent survenir – c'est-à-dire les contraintes associées au financement, aux délais, à la disponibilité des données et aux idées préconçues des clients – et montrons comment l'approche ESR peut aider à prendre ces contraintes en ligne de compte. Ce Chapitre pourra également s'avérer utile aux lecteurs qui découvrent le domaine de l'évaluation et pour qui il fera office d'introduction générale à la planification, la conception, la mise en œuvre, la diffusion des résultats et l'utilisation de toute évaluation quelle qu'elle soit.

Ce Chapitre a été conçu pour servir d'introduction à l'ESR et en même temps de résumé utile des points importants du présent ouvrage. Des références y sont faites aux autres Chapitres du livre *RealWorld Evaluation* où les divers points sont expliqués en détail.

Quelques-uns des tableaux et graphiques les plus importants des autres Chapitres sont reproduits dans l'Annexe à cette version détachable ; par exemple les Étapes de l'ESR (voir Tableau 1).

Cerner le champ de l'évaluation

Il est important que les personnes chargées de mener une évaluation comprennent bien les attentes de ceux qui demandent l'évaluation (les clients et les parties prenantes) – le contexte politique dans lequel s'inscrit le projet et l'évaluation. Il est également important de comprendre les décisions en matière de politiques et d'opérations sur lesquelles l'évaluation aura une incidence, ainsi que le degré de précision à atteindre lorsque l'on fournit l'information qui influera sur ces décisions (voir le Chapitre 2 du livre).

Comprendre les besoins du client

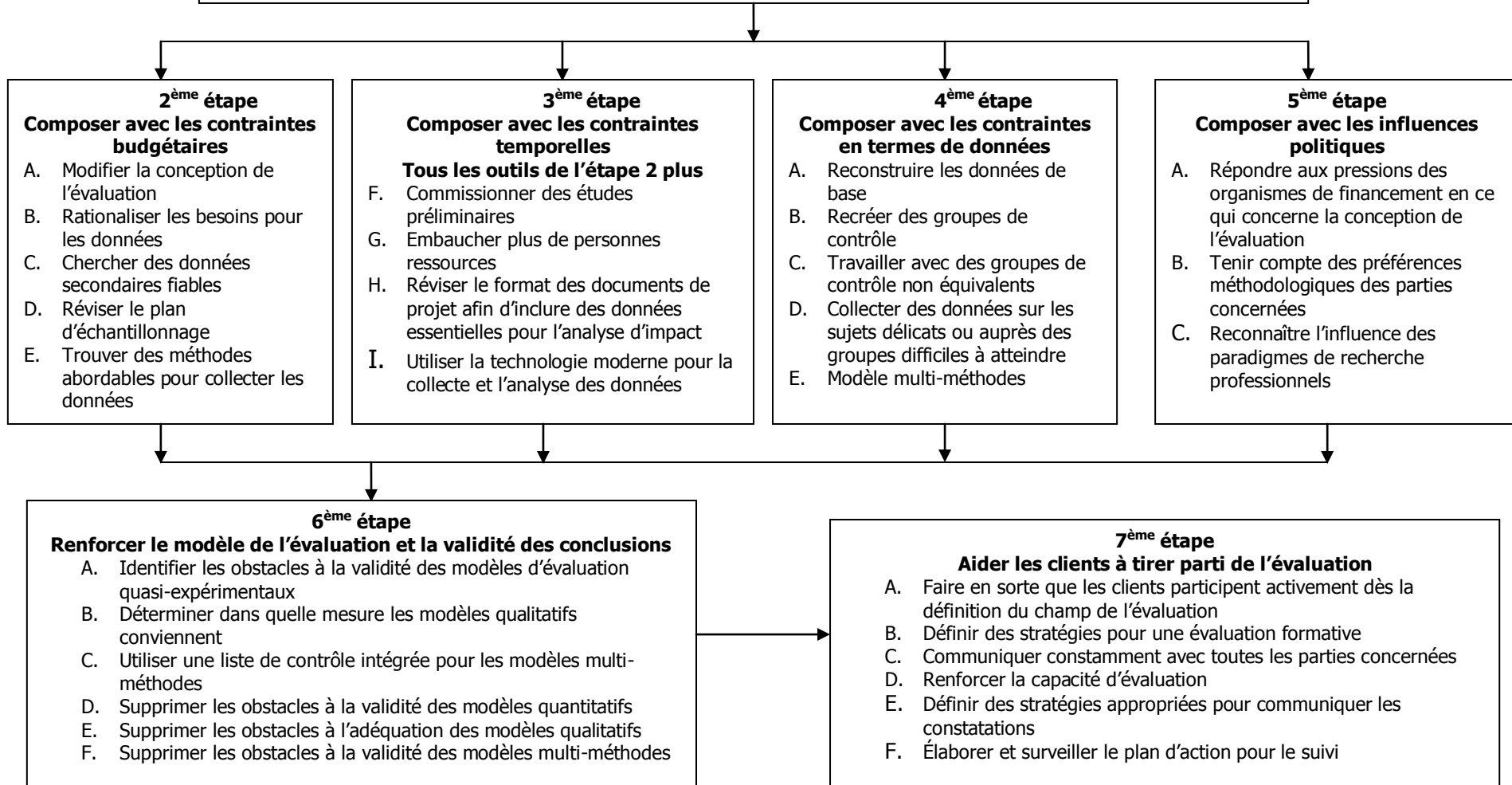
La première chose à faire absolument, lorsque l'on se prépare à faire une évaluation, et cela quel que soit le type d'évaluation envisagé, est de bien comprendre les priorités du client (l'organisme ou les organismes commanditant l'évaluation) et des principaux protagonistes (les personnes intéressées ou concernées par le projet) et de quelle information ils ont besoin. Le calendrier, les grands axes et le niveau de détail de l'évaluation doivent être déterminés en fonction des besoins du client en matière d'information et du type de décisions auxquelles doit servir l'évaluation.

Le travail nécessaire pour bien cerner les questions auxquelles il faut répondre peut aider ceux qui planifient l'évaluation à déterminer quelles collectes et analyses de données peuvent être omises parce qu'elles ne sont pas nécessaires. Il sera ainsi possible d'économiser de l'argent et de gagner du temps. L'évaluateur en situation réelle doit faire la distinction entre (a) l'information dont on a absolument besoin pour répondre aux grandes questions qui sont au cœur de l'évaluation et (b) les autres questions auxquelles il serait intéressant d'avoir des réponses, si l'on avait suffisamment de temps et de ressources, mais que l'on devra peut-être laisser tomber compte tenu des contraintes qui existent pour cette évaluation.

Tableau 1 L'approche de l'évaluation en situation réelle

1^{ère} étape : Planifier l'évaluation et en cerner le champ

- A. Définir le modèle de la théorie de programme
- B. Identifier les contraintes (temps, argent, contexte politique) dont il faudra tenir compte dans l'ESR
- C. Sélectionner l'approche qui répond le mieux aux besoins des clients compte tenu des contraintes pour l'ESR



Une autre décision à prendre en ce qui concerne les délais et les coûts concerne le choix des personnes qui devraient participer à la collecte des données et à l'examen des rapports d'évaluation. La philosophie de nombreux projets de développement consiste à promouvoir l'accroissement de l'autonomie des membres de la collectivité et des autres parties prenantes (administrateurs scolaires et enseignants, personnel des centres de soins de santé, organismes gouvernementaux locaux et organisations non gouvernementales, par exemple) et donc de les inviter à participer aux activités de suivi et d'évaluation. Lorsque l'on tient compte des contraintes dans le cadre d'une ESR, on pourra se poser la question suivante : « Dans quelle mesure est-ce important d'inclure des méthodes participatives et une représentation appropriée des parties prenantes dans l'évaluation? » Les méthodes participatives pour la collecte de données ont tendance à revenir plus cher et à prendre plus longtemps car il faut prévoir suffisamment de temps pour tisser des liens avec la collectivité et les autres protagonistes et instaurer des rapports de confiance avec eux. L'évaluateur en situation réelle, attentif aux coûts, doit déterminer si le client et les principaux protagonistes attachent une telle importance aux approches participatives qu'il faut prévoir le temps et le budget nécessaires pour adopter des approches participatives et *faire ça bien*.

Un autre défi rencontré lorsque l'on fait une ESR est que, du fait qu'il est souvent plus long et plus onéreux d'atteindre les groupes les plus démunis et les plus vulnérables, il n'est pas rare que les évaluateurs fassent l'objet de pression pour que ces groupes ne soient pas inclus dans les consultations. « Ce serait formidable de pouvoir consulter les squatteurs qui n'ont pas de titre légal de propriété mais, malheureusement, nous n'avons tout simplement assez d'argent et de temps. »

Une fonction capitale de la phase d'identification du champ de l'évaluation est qu'elle permet de comprendre si le manque de consultation des groupes concernés par le projet – y compris les groupes les plus démunis et les plus vulnérables – tient au manque de ressources ou au fait que le client ne considère pas leur participation comme une priorité. Souvent, le manque de temps et d'argent n'est qu'une excuse et il est donc important pour l'évaluateur de bien comprendre le point de vue du client avant de décider quelle approche choisir.

Comprendre le contexte politique

Font partie du contexte politique les priorités et les points de vue du client et des autres parties prenantes, la répartition du pouvoir et les relations entre ces derniers et les principaux protagonistes au sein du projet sur lequel porte l'évaluation et même les partis pris et les préférences philosophiques et méthodologiques de ceux qui effectuent l'évaluation (voir le Chapitre 6).

Il est important de ne pas partir des principes que toute influence politique est mauvaise et que les évaluateurs devraient pouvoir mener leur évaluation comme ils le jugent bon, sans aucune interférence des personnalités politiques et autres protagonistes « à l'esprit étroit » qui veulent s'assurer que leurs préoccupations soient prises en ligne de compte dans l'évaluation. Le but de l'évaluation est précisément d'aider à comprendre les politiques et les programmes sur lesquels les gens ont des opinions très marquées et souvent divergentes. Si une évaluation ne fait l'objet d'aucune pression ou influence politique, cela veut probablement dire que la question à l'étude est sans conséquence pour qui que ce soit, ou que l'évaluation est conçue de telle façon que les groupes concernés ne sont pas en mesure d'exprimer leurs opinions. Les évaluateurs ne devraient jamais assumer qu'ils ont raison et que les parties concernées, qui ne partagent pas leurs opinions sur les questions qui sont les plus importantes, la méthodologie qui convient ou la façon d'interpréter les constatations, ont des préjugés, sont mal informées ou ont tout simplement tort. Dans la **Liste de contrôle**

intégrée pour évaluer l'utilité et la validité des modèles qualitatifs, quantitatifs et multi-méthodes (Voir Annexe), le critère d'évaluation C porte sur la validité interne et l'authenticité des constatations de l'évaluation : « Les constatations sont-elles crédibles pour les gens sur qui porte l'étude et pour les lecteurs et a-t-on un portrait fidèle de ce que l'on étudie ? »

Si les principaux groupes concernés ne jugent pas l'analyse crédible, l'évaluateur devrait revenir en arrière et vérifier soigneusement la méthodologie et les hypothèses de base. Il ne peut se contenter de soupirer en se disant que c'est décidément bien difficile d'arriver à faire « comprendre » les constatations et les recommandations au client.

L'une des dimensions de l'analyse contextuelle qui a servi à élaborer le modèle de programme théorique (voir la section suivante) est l'examen de l'influence des facteurs politiques. Beaucoup des dimensions contextuelles – économiques, institutionnelles, environnementales et socioculturelles – influent sur la façon dont les groupes animés par des motivations politiques voient un projet et son évaluation. Il est essentiel de comprendre parfaitement ces facteurs contextuels si l'on veut comprendre les attitudes des principaux protagonistes concernés par le programme et l'évaluation de ce dernier. Une fois ces préoccupations comprises, il devient souvent plus facile de trouver comment faire face aux pressions exercées par les protagonistes pour influencer l'évaluation.

Le fait que de nombreuses évaluations de programmes soient commanditées par des gens animés par des motivations politiques n'a rien de surprenant. Certains clients qui souhaitent se servir de l'évaluation pour obtenir davantage de soutien pour le programme pourront ainsi s'opposer à l'inclusion de constatations ne le montrant pas sous son meilleur jour. D'un autre côté, le but réel mais non révélé du client qui a commandité l'évaluation était peut-être de réunir des pièces à conviction pour pouvoir renvoyer un gestionnaire, mettre fin à un projet, ou fermer un service. Lorsque tel est l'objet d'une évaluation, il est très rare qu'il soit formulé explicitement. Il se peut en outre que les protagonistes aient des opinions très divergentes sur un programme, son exécution, ses motivations, ses dirigeants et la façon dont il devrait être évalué. Les gens qui s'opposent à l'évaluation en cours arriveront parfois à l'empêcher ou à bloquer l'accès aux données, l'acceptation des résultats de l'évaluation ou la continuation d'un contrat d'évaluation.

Avant que l'évaluation ne débute, l'évaluateur devrait anticiper les divers types de problèmes politiques qui pourraient se présenter et y réfléchir, directement ou indirectement, avec le client et les principaux protagonistes. Le tableau 6.1 du Chapitre 6 montre certaines des façons dont le contexte politique peut affecter la conception et la mise en œuvre d'une évaluation et la façon dont les résultats sont communiqués et utilisés.

Les dimensions politiques ne concernent pas seulement les clients et les parties concernées. En font aussi partie les évaluateurs qui ont des approches qu'ils préfèrent parce qu'elles correspondent mieux à leurs points de vue personnels et professionnels sur ce qui constitue une pratique efficace et adaptée. Les différents évaluateurs, même ceux qui ont choisi de travailler ensemble sur un projet, prennent parfois des positions divergentes quant à leurs responsabilités publiques et morales. Comme tout le monde, les évaluateurs ont des valeurs qui leur sont propres. Il peut cependant être plus facile, pour de nombreux évaluateurs, de concevoir leur travail non pas comme l'imposition des valeurs de l'évaluateur mais plutôt comme un jugement objectif fondé sur les données en ce qui concerne le mérite du programme, ses faiblesses, son efficacité, son efficacité et la mesure dans laquelle il atteint son objectif. Les évaluateurs doivent être conscients du point de vue qui est le leur (et de leurs partis pris) et faire en sorte qu'il en soit tenu compte (voir la section A de la liste de contrôle intégrée à l'Annexe).

Les clients choisissent parfois les évaluateurs pour leur réputation d'honnêteté inébranlable, comptant sur cette réputation pour assurer la crédibilité des constatations et leur acceptation. Dans d'autres cas, le choix de l'évaluateur repose sur la coïncidence entre

les positions idéologiques prises par l'évaluateur et celles du client. Ces décisions peuvent être prises si discrètement qu'elles passent d'abord inaperçues lors des négociations amicales et des déclarations enthousiastes sur l'importance stratégique de l'évaluation envisagée. On trouvera au Chapitre 6 une discussion sur les options qui s'offrent à l'évaluateur lorsque les pressions exercées par les clients sont jugées inacceptables d'un point de vue éthique ou professionnel.

Les évaluateurs devraient également être conscients du fait que les orientations politiques des clients et des protagonistes peuvent influencer la façon dont les constatations de l'évaluation sont diffusées et utilisées. Il arrive que les clients ignorent les constatations qui ne leur plaisent pas ou bloquent leur diffusion en ne communiquant les rapports qu'à des lecteurs triés sur le volet, en ne distribuant que des résumés très brefs et dilués de ces rapports et en se chargeant de la présentation des rapports aux conseils d'administration et aux organismes de financement, s'acquittant ensuite de cette responsabilité de façon manipulatrice. On a vu des clients présenter des exposés et même des témoignages déformant les constatations de l'évaluation, entreprendre des activités de suivi qui ne sont pas suggérées – voire déconseillées – dans les rapports d'évaluation, et discréditer des évaluations et des évaluateurs qui représentent des menaces pour leurs programmes et leur prestige.

Un évaluateur sage sera conscient de ces facteurs et prêt à y faire face comme il convient. On trouvera au Chapitre 6 des suggestions de stratégies d'ESR pour faire face à des contraintes politiques de ce type et d'autres types lors de la conception de l'évaluation, de sa mise en œuvre et de la présentation et de l'utilisation des constatations.

Définir la théorie du programme

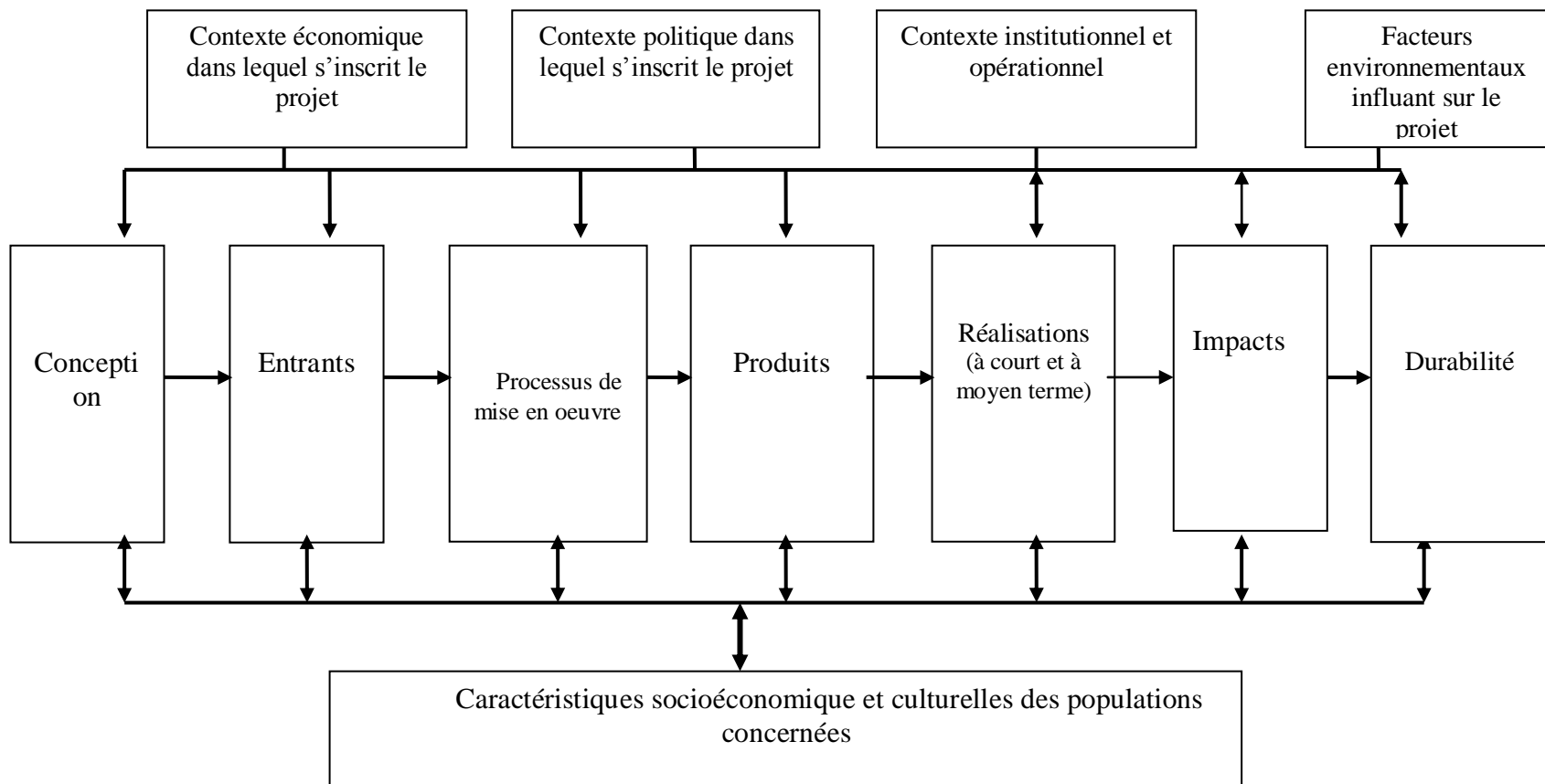
Avant de pouvoir procéder à une évaluation, il est nécessaire d'identifier la théorie ou le modèle logique, explicite ou implicite, qui sous-tend la façon dont le projet a été conçu (voir le Chapitre 9). Une fonction importante de l'évaluation d'impact est qu'elle permet de tester l'hypothèse selon laquelle les activités et les extraits du projet ont contribué aux effets souhaités, lesquels, combinés aux facteurs externes sur lesquels l'on comptait pour le projet, devaient avoir une incidence durable.

Définir la théorie du programme ou son modèle logique et une bonne pratique pour n'importe quelle évaluation. C'est un exercice particulièrement utile dans le cas des ESR, en ce sens que, du fait du budget, du temps disponible et des autres contraintes, il est nécessaire de déterminer sur quoi l'évaluation doit se concentrer en priorité. Un examen initial de ce que le projet a permis d'accomplir, au vu de son modèle logique, révélera dans certains cas de quelle information ou données on a besoin pour vérifier si la logique était solide et si le projet a permis d'accomplir ce qu'il fallait pour obtenir l'impact souhaité.

Quand le modèle logique a été clairement formulé dans le plan de projet, il peut servir à orienter l'évaluation. Lorsque ce n'est pas le cas, l'évaluateur doit le reconstituer en examinant les documents projet et au moyen de discussions avec l'organisme chargé de la mise en œuvre du projet, les participants au projet et les autres parties concernées (voir le Chapitre 9). Le plus souvent, cela nécessite un processus itératif au sein duquel la conception du modèle logique évolue au fur et à mesure que l'on apprend des choses au cours de l'évaluation.

En plus d'explicitement la théorie de cause à effet interne à partir de laquelle on a conçu le projet, un modèle logique devrait également identifier les caractéristiques socioéconomiques des populations affectées, ainsi que les facteurs contextuels comme les conditions économiques, politiques, organisationnelles, psychologiques et environnementales qui ont une incidence sur le groupe cible. On peut résumer comme suit les phases, ou niveaux, d'un modèle logique simple (voir la figure 1 et le Chapitre 2) :

Figure 1 Modèle théorique d'un programme simple



1. **Conception.** Comment le projet a-t-il été conçu? Qui l'a conçu? A-t-il été conçu juste par quelques membres de l'organisme donateur ou de celui chargé de la mise en oeuvre ou bien par un consultant externe? Ou bien est-ce que divers protagonistes, y compris les bénéficiaires visés, ont participé à sa conception? La conception a-t-elle été influencée par les enseignements tirés des évaluations de projets antérieurs qui utilisaient des approches similaires dans des conditions similaires?
2. **Ressources¹.** On entend par ressources les moyens financiers, humains, matériels et technologiques, ainsi que l'information disponible, utilisés pour l'action de développement.
3. **Processus de mise en oeuvre.** Ceci comprend les mesures prises ou le travail entrepris dans le cadre desquels les ressources – les fonds, l'assistance technique et les autres types de ressources – sont mobilisées pour parvenir à des extrants spécifiques. L'un des facteurs critiques est l'inclusion, ou non, des bénéficiaires visés et des autres parties prenantes dans le processus de mise en oeuvre et la façon dont on les inclut.
4. **Extrants.** Ils comprennent les produits et les services qui résultent lorsque l'action de développement a été menée à bien. On notera que les responsables de la mise en oeuvre du projet contrôlent directement les extrants – contrairement aux facteurs conceptuels externes qui pourraient affecter la livraison des extrants dans les délais convenus, ou leur qualité.
5. **Réalisations.** Il s'agit de ce que l'action vise à accomplir, ou a accompli, à court ou à moyen terme. On notera que les responsables de la mise en oeuvre du projet n'exercent pas un contrôle direct sur les réalisations. Celles-ci sont ce que les autres font eux-mêmes, sous l'influence, il est vrai, des extrants du projet.
6. **Impacts.** Ce sont les effets à long terme sur une population identifiable – positifs et négatifs, induits par une action de développement, directement ou non, intentionnellement ou non. Il peut s'agir d'impacts économiques, socioculturels, institutionnels, environnementaux ou technologiques, par exemple.
7. **Durabilité.** On entend par durabilité (ou viabilité) la continuation des bénéfices résultant d'une action de développement après la fin d'une intervention de grande envergure; Il s'agit de la résilience des avantages nets face au risque à travers le temps. Beaucoup de gens ne pensent même pas à cet aspect des choses ou tiennent pour acquis que les impacts seront durables. La durabilité des impacts n'est cependant pas garantie pour un certain nombre de raisons : il se peut que le projet reçoive des subventions qui s'arrêteront un jour; les conditions externes, comme les conditions météorologiques ou la situation politique ou économique, peuvent changer et menacer la stabilité des réalisations du projet; il est possible que les causes sous-jacentes du problème n'aient pas été traitées, au quel cas les actions dans le cadre du projet ne résoudre pas ces problèmes et n'auront pas des effets durables; enfin, il arrive que le projet introduise des techniques ou des technologies que les gens ou les organismes sont incapables de mettre en oeuvre eux-mêmes.

¹ Beaucoup de ces définitions sont tirées du glossaire de l'OCDE (Organisation de coopération et de développement économiques, 2002).

Chaque projet est conçu et mis en œuvre au sein d'une *situation* ou d'un *contexte* qui lui est propre et qui comprend les facteurs économiques, politiques, institutionnels et environnementaux, tant locaux que régionaux, ainsi que les caractéristiques socioculturelles des collectivités ou des groupes concernés par le projet. La théorie de programme doit incorporer tous ces facteurs par le biais d'une *analyse contextuelle*. Dans les cas où un projet est mis en œuvre à divers endroits, la performance et les réalisations varieront souvent de façon considérable entre les différents endroits du fait de la configuration différente des variables contextuelles.

Adapter les plans pour l'évaluation

Lorsque l'on commissionne une évaluation, il faut envisager un certain nombre de facteurs qui devraient être inclus dans le cadre de référence. La série de questions ci-dessous pourra être utile au client et aux évaluateurs embauchés pour effectuer l'évaluation. Elle pourra les aider à s'assurer que ces facteurs entrent en ligne de compte lorsque l'on planifie l'évaluation. Les réponses à ces questions pourront les aider à se concentrer sur les enjeux importants que l'on doit aborder lors de l'évaluation et à trouver des façons de composer avec les contraintes associées à une ESR.

- Qui a réclamé l'évaluation? Qui sont les principales parties prenantes? Ces gens ont-ils des idées préconçues quant au but de l'évaluation et aux constatations escomptées?
- Qui devrait participer à la planification et à la mise en œuvre de l'évaluation?
- Quelles sont les grandes questions auxquelles on doit apporter une réponse?
- S'agira-t-il d'une évaluation formative ou récapitulative? Est-elle principalement axée sur l'apprentissage et l'amélioration, sur la responsabilisation, ou un ensemble des deux?
- Y aura-t-il une autre phase ensuite, ou d'autres projets seront-ils conçus en fonction des constatations de cette évaluation?
- Quelles décisions seront-elles prises en réponse aux constatations de cette évaluation? Par qui?
- Quel niveau de rigueur s'impose pour pouvoir informer ces décisions?
- Quel est le champ/l'échelle de cette évaluation?
- De combien de temps aura-t-on besoin? Combien de temps aura-t-on?
- De quelles ressources financières aura-t-on besoin? Lesquelles aura-t-on?
- De quel type d'évaluation aurait-on besoin? Quel type est possible compte tenu des circonstances?
- L'évaluation devrait-elle s'appuyer principalement sur des méthodes quantitatives (QUANT), qualitatives (QUAL) ou une combinaison des deux?
- Devrait-on utiliser des méthodes participatives?
- Est-il possible/souhaitable de faire une enquête auprès des gens, des ménages ou d'autres entités?
- Qui devrait-on interviewer?
- Quel type d'échantillon (composition et taille) est nécessaire/possible?
- Avec quel type d'analyse répondra-t-on le mieux aux questions clés?
- À quels publics destine-t-on le ou les rapport(s)? Comment les constatations seront-elles communiquées à chacun des publics?
- Quel type de suivi est-il prévu? Par qui?

Choisir le meilleur type d'évaluation parmi les choix possibles

On trouvera au Chapitre 10, dans le tableau 10.1, un résumé de sept modèles d'ESR couramment utilisés (*voir l'Annexe*). Les deux premiers sont les plus rigoureux d'un point de vue méthodologique (la validité des conclusions est la moins attaquable) et sont recommandés lorsqu'il n'y a pas de contraintes majeures au niveau du budget et des délais et lorsqu'il est possible de commencer l'évaluation au début du projet. Les modèles 3 à 6 sont tous plus limités sur le plan méthodologique. Ils peuvent néanmoins produire des estimations raisonnablement crédibles des impacts d'un programme. Ils sont également plus souples en ce sens qu'il n'est pas nécessaire de commencer l'évaluation au début du projet. Le modèle 7 n'entre en jeu que lorsque le projet est presque terminé et ne couvre que le groupe participant au projet. C'est le plus faible sur le plan méthodologique. Néanmoins, étant données les contraintes qui régissent la conduite de nombreuses évaluations, c'est là la situation à laquelle doivent faire face de nombreux évaluateurs, malheureusement. S'il est utilisé de façon créative – en se servant d'approches multi-méthodes mettant à profit tous les modèles disponibles (quantitatifs et qualitatifs, collecte de données et approches en matière d'analyse – voir le Chapitre 14) – le modèle 7 peut fournir des estimations utiles sur le plan opérationnel de la mesure dans laquelle le projet contribue à l'obtention des effets souhaités. Étant donné les nombreux obstacles à la validité des conclusions, cependant, il est important d'examiner attentivement les limites inhérentes aux types de conclusions que l'on peut tirer d'une telle analyse. On trouvera au Chapitre 10 une description de chacun des modèles d'évaluation ainsi qu'une étude de cas illustrant la façon dont le modèle a été mis en application sur le terrain et une présentation des principaux obstacles à la validité et à l'adéquation des conclusions. Le tableau 10.4 résume les conditions dans lesquelles chacun des modèles marche et ne marche pas. Il est recommandé de consulter ce tableau très tôt dans le processus de planification de l'évaluation car cela aidera à éliminer un certain nombre d'options.

On trouvera ci-dessous une description de certaines des stratégies qui peuvent être utilisées pour renforcer la plupart des ESR. Les évaluateurs sont vivement encouragés à envisager l'utilisation de certaines de ces stratégies lorsque possible et souhaitable. Un certain nombre de facteurs déterminent le choix du modèle d'ESR qui convient, parmi lesquels :

Quand l'évaluation a-t-elle commencé? Au début du projet, en cours de projet, ou bien après qu'il soit arrivé à son terme?

Quand l'évaluation finira-t-elle? Est-ce que ce sera une évaluation isolée, effectuée en cours de projet (le cas de figure le plus courant étant l'examen à mi-parcours)? L'évaluation se terminera-t-elle à peu près au même moment que le projet (évaluation et rapport de fin de projet)? Ou bien continuera-t-elle une fois le projet terminé (évaluation longitudinale ou ex post)?

Quel type de comparaison utilisera-t-on? Trois possibilités : (a) un modèle aléatoire dans lequel des personnes, des familles, des groupes ou des collectivités sont choisis de façon aléatoire, avec des groupes témoins; (b) un groupe de comparaison choisi pour correspondre le mieux possible aux caractéristiques du groupe de projet; ou (c) aucun groupe témoin ou de comparaison.

Le modèle inclut-il une évaluation du processus? Même lorsque l'évaluation a pour objectif principal de mesurer les changements durables dans les conditions de vie de la population cible, elle doit identifier ce qui a le plus vraisemblablement causé ces changements. Il est donc nécessaire

d'évaluer la qualité du processus de mise en oeuvre du projet et sa contribution plausible aux impacts mesurés.

A-t-on des préférences pour le recours à des approches quantitatives, qualitatives ou multi-méthodes? Voir les Chapitres 11, 12 et 13.

Simplifier la conception de l'évaluation

Si l'on utilise des approches d'ESR, c'est parce que le temps, les ressources et les données dont on dispose – parfois aussi le contexte politique – ne permettent pas d'utiliser des modèles plus rigoureux pour l'évaluation. Dans ces circonstances, l'évaluateur doit travailler avec le client pour se mettre d'accord sur ce que l'on peut faire pour que l'évaluation nécessite moins de ressources et de temps et pour que les besoins en matière de données soient eux aussi révisés à la baisse, tout en veillant à conserver un degré de précision convenable. Les diverses options pour travailler avec un budget serré et dans des délais eux aussi serrés sont discutées aux Chapitres 3 et 4, le Chapitre 5 proposant une description des utilisations les plus efficaces des données disponibles. Les étapes ci-dessous sont recommandées pour définir le modèle le meilleur et le plus acceptable compte tenu des contraintes qui existent :

Consacrer le temps qu'il faut pour arriver à parfaitement comprendre les besoins les plus importants du client – de quelle information il a le plus besoin et quelles sont les contraintes politiques et autres qui le limitent dans son travail.

Qu'est-ce que l'évaluation doit réellement apporter au client?
Est-il essentiel d'avoir une analyse quantitative rigoureuse pour assurer la crédibilité de l'évaluation, ou bien une analyse qualitative approfondie est-elle plus importante et plus crédible pour les clients?
Quelle est l'information minimum dont on a absolument besoin et quelles sont les échéances absolues pour la première ébauche et pour le rapport final?
Qui sont les parties concernées sur qui l'évaluation est axée et dont les opinions sont capitales?

Examiner ce qu'il est possible de faire pour réduire les coûts (Chapitre 3) et les délais (Chapitre 4) et pour renforcer la base de données existante (Chapitre 5).

À partir de ces options, préparer plusieurs scénarios possibles pour atteindre les objectifs de l'évaluation, compte tenu des contraintes au niveau des ressources. Passer en revue la « Liste de contrôle intégrée pour évaluer l'utilité et la validité des modèles qualitatifs, quantitatifs et multi-méthodes » (Annexe) et évaluer les forces et les faiblesses de chacune des options du point de vue du client (ou des clients) et des principales parties concernées.

Si aucun des scénarios préparés ne permet de répondre aux besoins du client préparer deux autres scénarios :

Scénario 1 : Estimer l'argent ou le temps supplémentaire dont on aurait besoin pour répondre aux besoins du client (25 000 \$ de plus pour inclure un groupe de comparaison dans l'évaluation ou bien une extension de trois mois pour la soumission du rapport préliminaire, par exemple).

Scénario 2 : Indiquer les modifications qu'il faudrait apporter à l'approche proposée pour l'évaluation pour pouvoir la mener à bien avec les ressources disponibles.

Lorsque l'on prépare le scénario 2, il sera parfois possible de produire une estimation globale de l'impact du projet au niveau national mais pas de produire des estimations non regroupées de l'impact de diverses combinaisons de services, ou bien de l'impact pour différents sites du projet ou pour différents groupes socioéconomiques (hommes et femmes, salariés et travailleurs indépendants, par exemple, ou encore différents groupes ethniques).

Il est aussi possible qu'il faille baisser le niveau de fiabilité statistique (voir le Chapitre 14) pour pouvoir évaluer les différences statistiquement significatives entre le projet et les groupes de comparaison, optant pour un niveau de fiabilité de 0,10 (10 %) au lieu du niveau habituel de 0,05 (5 %).

Il faudra peut-être obtenir l'information au moyen de groupes de discussion dirigés (GDD) et de méthodes EPRM (bien qu'il ait au départ signifié évaluation participative en milieu rural, le terme désigne maintenant toute une gamme de méthodes d'évaluation participatives – on pense entre autres aux méthodes accélérées de recherche participative ou MARP) au lieu d'une enquête à domicile.

Toutes les options devraient être discutées avec le client, ainsi que les implications de chacune d'entre elles en termes du niveau de précision, des types d'analyse que l'on peut faire et de la crédibilité des constatations pour les diverses parties concernées. Il est essentiel de veiller à ce que le client comprenne bien ces options et les concessions qu'elles représentent avant de décider comment procéder. Dans certains cas, le client demandera à l'évaluateur « Quelle approche est la meilleure à votre avis, puisque c'est vous le spécialiste? » Si on lui pose cette question, l'évaluateur doit répondre que c'est une décision stratégique qui doit être prise en consultation avec le client, le rôle de l'évaluateur étant simplement de formuler des avis quant aux implications techniques de chacune des options en se limitant à la précision, aux types d'analyse et à la crédibilité sur le plan professionnel.

Identifier les types d'analyse et de comparaison qui sont essentiels pour l'évaluation

Un facteur clé dans le choix d'un modèle d'évaluation, et pour déterminer la taille et la structure de l'échantillon, est le type d'analyse et les niveaux de précision statistique dont on a besoin. Il sera bon d'envisager trois types d'évaluation :

1. *Les évaluations préliminaires ou sous forme de recherche* qui ont pour objet de déterminer si le concept et l'approche de base pour le projet « marchent ». C'est souvent l'approche choisie lorsqu'un nouveau type de service est mis à l'essai ou lorsqu'un service existant doit être fourni d'une nouvelle façon, ou bien encore quand on veut atteindre de nouveaux groupes cibles. Seront par exemple incluses dans les questions clés de l'évaluation des questions comme :
 - a. Les fermiers sont-ils disposés à expérimenter en utilisant de nouvelles sortes de graines?

- b. Les nouvelles méthodes d'enseignement sont-elles bien reçues par les écoles, les élèves et les parents et y a-t-il des preuves initiales de l'amélioration de la performance?
- c. Si les femmes pauvres se voient accorder des prêts, sont-elles en mesure de se servir de l'argent pour créer ou agrandir une petite entreprise?
- d. Quels groupes bénéficient-ils le plus et le moins, et pourquoi?

2. *Les évaluations quasi expérimentales à petite échelle ou les modèles qualitatifs* pour déterminer s'il existe des preuves que le projet a des retombées significatives pour la population cible. Certains modèles incluent un groupe de comparaison, alors que d'autres ont recours à une comparaison plus générale avec des collectivités similaires au moyen de techniques d'EPRM ou de groupes de discussion. On traite des questions d'attribution (quelle aurait été la situation du groupe de projet si celui-ci n'avait pas eu lieu?) mais avec moins de rigueur que dans le cas des études d'impact à grande échelle (voir ci-dessous). On trouvera parmi les questions clés des questions comme :

- a. La situation des bénéficiaires visés du projet (les personnes, les familles, les écoles ou les collectivités, par exemple) s'est-elle améliorée grâce au projet?
- b. Dans quelle mesure est-on sûr que les changements constatés sont le résultat du projet et non de facteurs extérieurs tels qu'une reprise de l'économie locale?
- c. Est-il probable que le projet aura des effets similaires si on le reproduit ailleurs? Où marcherait-il bien, ou moins bien, et pourquoi?
- d. Quels facteurs contextuels (économiques, politiques, institutionnels, environnementaux et culturels) ont une incidence sur la réussite du projet? (voir le Chapitre 9 pour une discussion des facteurs contextuels et des variables de médiation).
- e. À qui le projet a-t-il profité et à qui n'a-t-il pas profité?

3. *Les évaluations à grande échelle des impacts* pour déterminer, avec plus de rigueur statistique, l'ampleur de l'effet produit (défini numériquement en termes de pourcentage ou de changement quantitatif) et à qui le projet a profité ou n'a pas profité. Dans l'idéal, on se servirait pour l'évaluation d'une approche multi-méthodes intégrant des méthodes quantitatives et qualitatives. On pourra inclure dans les questions clés les questions suivantes :

- a. Quels impacts quantitatifs (effets durables de haut niveau) le projet a-t-il eus? On s'intéresse principalement au « combien », pas seulement au « quoi ».
- b. Quelle est la qualité des services (par rapport aux autres programmes, aux normes en vigueur et dans l'opinion des groupes cibles)?
- c. Les effets du programme sont-ils statistiquement significatifs, « hors de tout doute raisonnable »?
- d. Qui en a le plus et le moins profité et y a-t-il des groupes qui n'en ont tiré aucun avantage ou dont la situation a empiré?
- e. Quelles sont les variables susceptibles de jouer un rôle (caractéristiques socioéconomiques des groupes de projet et facteurs culturels ayant une incidence sur la participation, par exemple) et d'influer sur la magnitude des impacts?

Pour les *évaluations préliminaires*, il suffirait sans doute d'une analyse descriptive utilisant des techniques comme l'observation, des interviews avec quelques membres

choisis de la population cible, des interviews des principaux répondants et peut-être des groupes de discussion. Une enquête simple et rapide pourrait également être faite pour collecter l'information de base sur la population visée par le projet. Il ne serait vraisemblablement pas nécessaire d'avoir un groupe de comparaison formel – on pourrait visiter des collectivités ou des régions similaires pour déterminer dans quelle mesure elles sont similaires ou différentes de celle sur laquelle porte le projet.

Pour les *évaluations quantitatives ou qualitatives à petite échelle des effets directs et des impacts*, il est utile, bien que pas toujours possible, d'identifier un groupe de comparaison pour aider à estimer quelle aurait été la situation du groupe de projet si ce dernier n'avait pas eu lieu (le scénario en l'absence d'intervention). Dans l'idéal, une approche multi-méthodes sera utilisée pour évaluer la qualité et la quantité des services et des impacts. Le modèle est en outre beaucoup plus rigoureux si l'on parvient à obtenir des données de base (sous quelque forme que ce soit). De simples comparaisons statistiques, comme les écarts dans les moyennes ou les proportions, devraient être effectuées entre le projet et les groupes de comparaison.

Pour les *évaluations à grande échelle des impacts*, des échantillons relativement grands (nécessitant des centaines d'observations tant dans le groupe de projet que dans les groupes de comparaison) sont nécessaires pour que l'on puisse faire une analyse multivariable pour contrôler statistiquement les différences entre le groupe de projet et les groupes de comparaison et estimer l'influence quantitative des variables susceptibles d'intervenir. Ici encore, une approche multi-méthodes devrait être utilisée afin que les estimations quantitatives soient complétées par des descriptions qualitatives du contexte du projet, du processus de mise en oeuvre du projet, de la qualité des services et des opinions et des expériences des bénéficiaires et des organismes et du personnel chargés de la mise en oeuvre du projet.

Évaluer les obstacles à la validité et l'adéquation des différents modèles

L'approche ESR permet d'évaluer les forces et les faiblesses des différentes phases du modèle d'évaluation dans le but d'identifier les facteurs qui déterminent la validité des conclusions et des recommandations. Cela est important pour n'importe quelle évaluation, mais encore plus pour l'ESR, les méthodologies traditionnelles ayant souvent été simplifiées du fait de contraintes temporelles et budgétaires, ou parce que certaines des données dont on a besoin ne sont pas disponibles. Plusieurs facteurs influent sur l'adéquation du modèle choisi pour l'évaluation et de ses constatations (voir le Chapitre 7, tableau 7.2), dont les facteurs suivants :

- L'adéquation de l'axe, l'approche et les méthodes d'obtention des types d'information choisis
- La disponibilité des données et des sources d'information
- La mesure dans laquelle les données collectées appuient les interprétations quant à la performance et aux impacts du programme
- Les compétences de l'équipe d'évaluation, en ce qui concerne la méthodologie d'évaluation et les domaines spécifiques du programme

Pour les évaluations quantitatives, quatre groupes d'obstacles à la validité identifiés par Cook et Campbell dans les années 1960 (voir Shadish, Cook, et Campbell, 2002, pour une version mise à jour) sont généralement acceptés. On trouvera dans l'Annexe (sections F, G, H et I de la Liste de contrôle) une description et une explication des différents types d'obstacles en ce qui concerne les modèles d'évaluation quantitative qui appartiennent aux quatre catégories décrites ci-dessous.

Les obstacles à la validité statistique des conclusions (voir l'Annexe, section F de la Liste de contrôle) – Inférences/conclusions inexactes quant aux effets des interventions de projet sur les effets directs et les impacts : Les problèmes peuvent venir de la mauvaise application d'un test statistique ou de l'insuffisance de l'échantillon (si celui-ci n'est pas représentatif de toute la population, par exemple). Un autre problème fréquent est la conclusion, au terme de l'analyse, que le projet n'a pas eu un effet statistiquement significatif, alors qu'en fait l'échantillon était trop petit (souvent pour des raisons financières) pour que l'on puisse déceler cet effet, même s'il avait existé. Il faut souvent trouver un compromis entre la réduction de la taille de l'échantillon pour diminuer les coûts et la nécessité de s'assurer qu'il soit assez grand pour que l'on puisse détecter les effets si effets il y a eu. Qui plus est, on peut s'attendre à ce que beaucoup de projets bien conçus et bien exécutés ne produisent que de petits effets, ce qui rend plus difficile la détection de ces derniers. Les questions relatives aux erreurs de Type II (conclure à tort qu'il n'y a pas eu d'effet), la puissance du test, l'ampleur de l'effet et l'estimation de la taille de l'échantillon sont traitées au Chapitre 14.

Les obstacles à la validité interne (Section G de la liste de contrôle) – Inférences inexactes selon lesquelles les interventions de projet ont « causé » un effet direct ou un impact, ou y ont contribué : De nombreux problèmes de validité interne viennent de la façon dont les participants au projet sont choisis (ils sont différents, de façon significative, du groupe de comparaison), les caractéristiques du groupe de projet changent au fil du projet, parce que des gens abandonnent ou bien parce que certaines expériences pendant le projet changent la façon dont les gens réagissent.

Les obstacles à la validité de contenu (Section H de la Liste de contrôle) – La mesure dans laquelle des inférences peuvent légitimement être faites à partir des concepts théoriques (définition des concepts clés) sur lesquels repose la théorie du programme : beaucoup des concepts clés sont difficiles à définir (pauvreté, vulnérabilité, bien-être, milieu de travail hostile, par exemple) et plus difficiles encore à mesurer. Si ces concepts clés ne sont pas définis et mesurés avec suffisamment de précision et de clarté, il sera plus difficile pour les évaluateurs de comprendre et d'interpréter comment le projet a fonctionné et ce qui a été accompli.

Les obstacles à la validité externe (Section I de la Liste de contrôle) – Inférences inexactes quant à l'application possible des constatations de l'évaluation à d'autres gens, à un autre moment ou dans un autre endroit : La plupart des évaluations quantitatives ont pour objet de déterminer dans quelle mesure les constatations de l'évaluation peuvent être généralisées et appliqués à une population plus vaste (toutes les collectivités à faible revenu, toutes les travailleuses non qualifiées, tous les élèves du secondaire, par exemple). Les caractéristiques de la population du projet ne sont pas forcément typiques de la population générale, et ce de diverses manières, et la généralisation des constatations peut donc induire en erreur. Un programme d'alphabétisation pour les adultes peut par exemple avoir été une réussite, en partie au moins à cause de l'appui enthousiaste de la chambre de commerce locale qui a fourni gratuitement aux participants des moyens de transport, des cahiers et des collations. La réussite du programme ne permet donc pas de

conclure qu'il réussirait aussi dans d'autres villes où l'on n'a pas le même soutien au niveau local.

On peut se servir de la liste de contrôle pour les obstacles à la validité lorsque l'on veut identifier et évaluer les faiblesses potentielles des sept modèles d'ESR quantitatives les plus souvent utilisés (voir les Chapitres 7 et 10). Les lecteurs qui ne sont pas spécialistes de l'analyse statistique et de l'évaluation quantitative trouveront peut-être les sections F à I de la liste de contrôle intégrée difficiles à suivre. Cela vaut toutefois la peine de jeter un coup d'œil aux différentes catégories pour se faire une idée de toute la gamme de facteurs qui peuvent influencer sur la validité des modèles d'évaluation quantitative, et il est toujours possible de demander l'avis d'un spécialiste quand il faut concevoir des évaluations quantitatives rigoureuses et déterminer leur validité.

Pour les lecteurs qui ne sont pas spécialistes de la question, les sections A à E de la Liste de contrôle (voir ci-dessous) fournissent suffisamment de conseils pour déterminer la validité de la plupart des évaluations. Ces méthodes sont normalement utilisées pour déterminer la validité (on parle souvent d'adéquation) des modèles d'évaluation multi-méthodes et qualitatives mais elles complètent aussi les catégories F à I pour l'évaluation des modèles d'évaluation quantitative.

Preuves à l'appui et objectivité (Section A de la Liste de contrôle) : Les conclusions sont-elles tirées des preuves disponibles et la recherche est-elle relativement libre des préjugés de ses auteurs? Il y a cinq sous-catégories : si les méthodes et les procédures sont bien décrites, si des données sont présentées pour appuyer les conclusions, si la partialité des chercheurs est reconnue et prise en compte et si des hypothèses concurrentes ont été envisagées.

Fiabilité et constance (Section B de la Liste de contrôle) : Le processus utilisé pour l'étude est-il constant, cohérent et raisonnablement uniforme au fil du temps, d'un chercheur à l'autre et d'une méthode à l'autre? Il y a huit sous-catégories concernant la clarté des questions de recherche, l'application des méthodes de collecte des données et les façons de procéder pour vérifier et valider les méthodes. Alors que les évaluations quantitatives visent à garantir l'uniformité et la normalisation, de nombreuses évaluations qualitatives utilisent des *modèles émergents* où les méthodes de collecte des données et l'interprétation évoluent et changent au fur et à mesure que l'évaluation progresse. Le plus difficile est de déterminer si les processus émergents sont bien documentés et évalués pour que le lecteur puisse comprendre les preuves sur lesquelles reposaient les conclusions.

Validité interne, crédibilité et authenticité (Section C de la Liste de contrôle) : Les constatations sont-elles crédibles pour les personnes étudiées et les lecteurs et présentent-elles un portrait fidèle de ce qui a été étudié? Les huit sous-catégories portent sur la profondeur et la richesse de la collecte des données et de leur analyse, la perception qu'ont les différents groupes de la crédibilité de l'analyse, la conformité des constatations avec la théorie et les hypothèses existantes et la mesure dans laquelle les constatations sont bien soutenues par les données.

Validité externe, transférabilité et véracité dans d'autres contextes (Section D de la Liste de contrôle) : Les conclusions peuvent-elles être étendues à d'autres contextes et dans quelle mesure peut-on les généraliser? Les 11 sous-catégories permettent de déterminer jusqu'à quel point il est

possible de le faire et à quels autres groupes, ou quelles autres populations, les constatations peuvent être étendues.

Utilisation, application et accent mis sur l'action (Section E de la Liste de contrôle): De quelle utilité les constatations ont-elles été pour le client, les chercheurs et les collectivités étudiées? Les sept sous-catégories proposent des indicateurs pour déterminer l'utilité.

Quand et comment utiliser la liste de contrôle pour les obstacles à la validité et l'adéquation

On se référera à la liste de contrôle tout au long de l'évaluation pour s'assurer que tout progresse comme il faut et pour cerner et régler rapidement les problèmes qui pourraient affecter la validité de la méthodologie et des conclusions. C'est dans les moments suivants qu'il est le plus important de se servir de la liste de contrôle :

Lorsqu'on est en train de concevoir l'évaluation et de planifier les méthodes de collecte de données et d'analyse. On peut se servir de la liste de contrôle pour identifier les problèmes potentiels et envisager des façons de les régler. Cela est particulièrement utile dans le cadre des ESR pour déterminer si les méthodes proposées pour réduire les coûts et les délais et pour travailler avec des bases de données restreintes risquent de compromettre la validité de l'évaluation. Quels obstacles supplémentaires à la validité surgissent, par exemple, si l'on supprime le groupe de comparaison de référence ou si des données sur le revenu sont collectées au moyen de groupes de discussion au lieu d'enquêtes à domicile?

Lorsque la collecte de données est pratiquement terminée. La liste de données peut aider à repérer des obstacles qui ont pu surgir pendant la collecte de données (taux de non-réponse plus élevé qu'on ne s'y attendait ou confusion quant au concept de chômage, par exemple). On se servira de la liste de contrôle le plus rapidement possible après que la collecte de données soit terminée (ou même avant qu'elle soit finie) pour qu'il soit encore temps de prendre des mesures correctives.

Lorsque le rapport d'évaluation préliminaire est terminé. Dans l'idéal, il serait encore temps de prendre des mesures correctives, mais si ce n'est plus possible, on devrait inclure la liste de contrôle dans le rapport pour pouvoir identifier et décrire clairement les obstacles potentiels à la validité et indiquer comment ils pourraient affecter les conclusions et les recommandations.

L'Annexe du livre contient une feuille de travail pour identifier les obstacles à la validité et à l'adéquation dans les projets d'ESR et composer avec elles. Cette feuille de travail peut être utilisée à tous les stades de l'évaluation pour identifier les obstacles à la validité et à l'adéquation, déterminer dans quelle mesure chacun de ces obstacles compromet la raison d'être de l'évaluation et identifier les mesures que l'on peut prendre pour essayer de supprimer les principaux obstacles, ou au moins composer avec eux. On trouvera dans l'Annexe du livre un exemple de feuille de travail remplie qui montre comment on peut se servir de la feuille de travail pour évaluer une évaluation d'un projet de logement.

Déterminer quelles méthodologies conviennent

Les approches et les méthodes quantitatives et qualitatives sont, en règle générale, conçues à des fins différentes. Les forces et les faiblesses potentielles des deux approches sont présentées aux Chapitres 11 et 12. Compte tenu de ces forces et faiblesses, on considère dans l'approche d'ESR que le modèle d'évaluation le plus rigoureux et le plus fiable est probablement une combinaison d'approches quantitatives et qualitatives. Le Chapitre 13 propose une discussion détaillée des façons dont un modèle multi-méthodes combinant systématiquement des approches quantitatives et qualitatives peut être utile pour les ESR. Ce type d'approches multi-méthodes est particulièrement précieux pour les ESR car la combinaison des différentes méthodes de collecte et d'analyse de données peut permettre de repérer certains des obstacles à la validité suite aux compromis imposés par les contraintes en termes d'argent, de temps et de données. Elles permettent aussi de remédier à ces faiblesses, ou au moins de les minimiser. Lorsque l'on diminue la taille de l'échantillon, par exemple, les erreurs d'échantillonnage augmentent et il devient plus difficile de détecter des différences significatives. Si on utilise des techniques qualitatives, comme l'EPRM, pour obtenir des estimations indépendantes des impacts du projet, et si les résultats convergent avec l'analyse statistique des enquêtes par sondage, la confiance dans les constatations sera sans doute accrue.

Selon de nombreux auteurs, lorsque l'on utilise des approches multi-méthodes, les évaluations ont toujours une orientation quantitative ou qualitative prédominante (orientation théorique), l'autre approche étant utilisée en complément. L'orientation théorique sera déterminée par l'orientation professionnelle du chercheur ou par la préférence du client. D'autres auteurs suggèrent cependant une approche intégrée qui ne favorise aucune des deux approches. Que l'on adhère à l'une ou l'autre de ces positions, il demeure que certains programmes se prêtent plus naturellement aux méthodes d'évaluation quantitative (des programmes de très grande envergure concernant des milliers de gens, par exemple) alors que, dans d'autres cas, les méthodes qualitatives sembleront peut-être mieux convenir (pour un programme dont le but est d'améliorer la qualité des pratiques pédagogiques ou d'introduire la culture de légumes dans plusieurs villages d'assez petite taille, par exemple). Cependant, dans tous les cas, le choix des méthodes sera influencé par les préférences des évaluateurs et des clients.

Tout au long du présent ouvrage, nous encourageons vivement les évaluateurs à sélectionner les outils de collecte et d'analyse de données qui correspondent le mieux aux besoins du client et à la nature du programme à l'étude et à éviter de choisir certaines méthodes simplement parce qu'elles sont qualitatives ou quantitatives.

Moyens pour renforcer les modèles d'évaluation en situation réelle

Du fait des contextes dans lesquels les ESR sont mises en œuvre, un certain nombre d'obstacles à la validité ont de fortes chances d'affaiblir la qualité des données et la validité des conclusions lorsque les modèles quantitatifs conventionnels et quasi expérimentaux sont utilisés séparément. (C'est pour cette raison que l'utilisation de modèles multi-méthodes est vivement recommandée dans les ESR.) Ces questions sont traitées plus en détail au Chapitre 10. Voici quelques-unes des limites inhérentes aux modèles quantitatifs conventionnels :

Des problèmes peuvent surgir quant à la fiabilité de la mesure des indicateurs-clés, en particulier lorsque ceux-ci concernent des questions délicates comme l'utilisation de drogues illicites, le contrôle des

ressources, la violence familiale et la restriction de la mobilité ou des activités économiques des femmes par la société.

Il peut être difficile de saisir les variations au niveau de la mise en oeuvre du projet, de la qualité des services et de l'accès des différents groupes aux services et aux avantages.

Les modèles conventionnels n'analysent pas les facteurs contextuels qui ont une incidence sur les résultats et les impacts du projet aux différents endroits.

Les différences importantes (non-équivalence) entre le groupe de projet et le groupe de comparaison – en particulier celles qui sont difficiles à quantifier (motivation ou organisation de la collectivité, par exemple) – sont difficiles à capter.

On trouvera ci-dessous un certain nombre de façons de procéder pour renforcer ces modèles (voir aussi l'encadré 7.2 au Chapitre 7 pour des exemples des façons dont on peut tenter de remédier aux obstacles les plus communs à la validité statistique, interne, conceptuelle et externe). Les évaluateurs devraient envisager d'incorporer certaines des procédures suivantes dans les modèles d'ESR si besoin est.

Fonder l'évaluation sur un modèle de programme théorique

Comme nous l'avons vu un peu plus haut dans le présent chapitre, la formulation d'un modèle de programme théorique aide à cerner les principaux enjeux et les hypothèses auxquels on devrait consacrer les ressources restreintes dont on dispose pour l'évaluation (voir les Chapitres 2 et 9). On peut également utiliser un modèle théorique en complément d'un modèle quasi-expérimental et décrire les facteurs contextuels qui influent sur la mise en oeuvre et les effets directs. Cela peut en outre aider à interpréter les constatations de l'évaluation et à déterminer si le projet devrait être poursuivi ou reproduit.

Compléter l'évaluation quantitative (sommativ) par une évaluation des processus

Pour une **évaluation des processus**, on se sert de méthodes quantitatives et qualitatives, ainsi que d'approches multi-méthodes permettant d'observer et d'évaluer les processus de mise en oeuvre du projet et de faire des recommandations quant aux façons de renforcer les phases suivantes d'un programme en cours. On peut poser les questions suivantes :

Comment les différentes composantes du programme ont-elles été mises en oeuvre et dans quelle mesure la mise en oeuvre sur le terrain correspondait-elle fidèlement au plan de projet ou au manuel des opérations?

Pour les projets en cours, comment la qualité des services pourrait-elle être évaluée et améliorée? A-t-on des preuves que ces services ont les effets souhaités?

Qui a accès aux services et qui les utilise, et qui n'y a pas accès ou ne les utilise pas? Pourquoi certains groupes n'utilisent-ils pas les services?

Est-ce que le projet a été conçu et organisé de façon participative, par un petit groupe ou bien de façon descendante? Qui a participé à la prise des décisions pendant la mise en oeuvre?

Quelle proportion de la collectivité (bénéficiaires visés) est au courant du projet? Ces gens-là ont-ils des renseignements exacts? Que pensent-ils du projet?

Quels sont les rapports entre les organisateurs du projet et la collectivité?
Quels organismes, gouvernementaux ou autres, sont au courant de ce projet
et que pensent-ils de la qualité des services et de l'efficacité du projet?

Incorporer une analyse contextuelle

L'analyse contextuelle consiste à évaluer l'influence des facteurs économiques, politiques, organisationnels et environnementaux sur la mise en œuvre et les effets directs des projets. C'est ce qu'on appelle, dans les modèles de programme théoriques, les facteurs médiateurs (voir Chapitre 9). On se penche également sur l'influence des caractéristiques socioculturelles préexistantes des populations cibles sur la façon dont le projet est accueilli par les différents groupes.

L'analyse contextuelle est le plus souvent qualitative (entrevues avec des répondants clés, examen des documents du projet et observation-participation). Elle peut cependant inclure l'analyse quantitative des données des enquêtes à domicile. Les variables contextuelles peuvent par ailleurs être transformées en variables numériques (des « variables nominales », par exemple) et incorporées dans l'analyse multi variables (voir le Chapitre 13, note 1).

Reconstituer les conditions de base

Lorsque l'évaluation commence, le projet a souvent débuté depuis longtemps. Parfois même, il est presque fini. Il arrive souvent dans ces circonstances que l'on découvre qu'aucune donnée de base n'a été collectée au début du projet. Le plus souvent, c'est le cas pour le groupe de projet. L'absence de données de base étant généralement l'un des plus grands obstacles à la validité, l'ESR permet de *reconstituer* les données de base de plusieurs façons. Ces approches sont décrites au Chapitre 5, Tableau 5.1, et incluent :

- L'utilisation de données secondaires (voir la section suivante)
- L'utilisation de souvenirs personnels (on demande aux répondants de se souvenir de la situation de leur famille ou de leur collectivité à l'époque où le projet a commencé)
- L'utilisation de l'EPRM et autres techniques participatives pour *reconstituer* l'histoire de la collectivité et évaluer les changements entraînés par le projet.
- Des entrevues avec des répondants-clés, de préférence des gens qui connaissent la collectivité cible ainsi que d'autres collectivités et ont donc un sens des changements relatifs qui ont eu lieu au fil du temps.

Bien que toutes ces méthodes fournissent des renseignements potentiellement utiles, toute méthode recourant au souvenir comporte forcément des obstacles majeurs à la validité. Ceci est dû au manque de souvenirs précis, à la tendance à ne pas se souvenir clairement de la période concernée (de telle sorte que l'on pourra se souvenir d'événements qui ont eu lieu avant et les raconter comme s'ils avaient eu lieu après que le projet ait commencé et vice versa) et, dans certains cas, à une distorsion délibérée. Il est par conséquent important de se méfier de toutes les données provenant de souvenirs et de toujours se servir d'approches multi-méthodes pour trianguler des estimations indépendantes de l'information rapportée par diverses sources.

*Utiliser des données secondaires**Planification administrative et contrôle des données collectées par l'organisation évaluée.*

Les données collectées dans le cadre d'une évaluation diagnostique antérieure au projet et celles collectées durant le projet grâce au système de suivi de la personne ou de l'organisme qui le met en oeuvre sont des sources d'information importantes, et pourtant souvent négligées, pour reconstituer les données des conditions de base (voir le Chapitre 5). Pour la plupart des projets, on collecte beaucoup de données à des fins administratives et de suivi et ces recueils de données contiennent souvent de l'information qui peut être utile pour reconstituer l'information sur les conditions de la population de projet à l'époque où ce dernier a commencé (données de base). Le plus souvent, on recherchera des sources d'information comme :

- Les documents de planification et les études de faisabilité préparés avant que le projet ne commence

- Les descriptions des caractéristiques socioéconomiques des personnes et des familles qui ont fait une demande pour recevoir les services, ou qui les reçoivent

- La participation à des réunions communautaires et, le cas échéant, les rapports sur ces réunions (comptes rendus par exemple)

- Les rapports d'activité préparés par le personnel de l'organisme ou d'autres personnes participant à la mise en oeuvre du projet. Au pire, ces rapports fournissent des renseignements sur le processus. Dans l'idéal, il y est aussi fait mention des changements observés au niveau des connaissances, de l'attitude et des pratiques des clients.

Si des données administratives de ce type peuvent être extrêmement utiles en tant que données de base, il est important de se souvenir que les données n'ont pas été collectées pour servir à l'évaluation et qu'elles ne sont pas forcément idéales (il se peut qu'elles soient incomplètes ou mal organisées, ou encore qu'elles ne contiennent pas l'information dont on a besoin pour l'évaluation). On trouvera dans la deuxième section du Chapitre 5 une liste de questions que l'on devrait poser lorsque l'on évalue la qualité et l'utilité de cette information pour la reconstitution des données de base.

Annales tenues à l'extérieur de l'organisation. Les annales d'autres programmes ou projets dans la même région peuvent souvent fournir des renseignements sur les conditions avant que le projet ne commence. Ainsi, des sondages sont souvent faits pour estimer le nombre d'enfants qui ne sont pas scolarisés, les sources et les coûts pour l'approvisionnement en eau, ou encore la disponibilité de microcrédits. Des données statistiques plus générales peuvent aussi être disponibles, comme les taux de scolarisation, la mortalité infantile, les prix des produits agricoles, les prêts accordés dans le cadre de microcrédits et les schémas de transport. Il est important d'évaluer les forces et les faiblesses de ces annales et en particulier les points suivants :

- Le temps qui sépare le début du projet (moment auquel on a besoin de données sur les conditions de base) et le moment auquel on collecte les données secondaires. Les écarts temporels sont particulièrement importants dans les cas où les conditions économiques générales ont changé entre la date du sondage et le lancement du projet.

- Les différences dans la population couverte. Tenait-on compte dans les sondages, par exemple, des emplois dans le secteur non structuré, et pas

seulement dans le secteur structuré? Les piétons étaient-ils inclus avec les modes de transport véhiculaires?

De l'information a-t-elle été collectée sur les variables clés du projet et sur les impacts potentiels? Les données secondaires sont-elles statistiquement valides pour la population cible spécifique du projet que l'on est en train d'évaluer?

Est-ce que l'information est disponible pour les hommes et pour les femmes? Ou bien toute l'information a-t-elle été obtenue auprès d'une seule personne (en général le « chef du ménage ») et rassemblée pour tous les membres du ménage (voir encadré 5.2)?

Utiliser des approches multi-méthodes pour renforcer la validité des indicateurs et améliorer l'interprétation des constatations

Il est possible de renforcer les sept modèles d'évaluation décrits au Chapitre 10 en utilisant des modèles multi-méthodes (voir Chapitre 13) qui combinent des approches quantitatives et qualitatives de l'une des façons suivantes :

Études préliminaires pour comprendre le contexte et identifier les principaux enjeux et les hypothèses à vérifier. Ceci est particulièrement important pour la construction du modèle de programme théorique (voir le Chapitre 9)

Analyse de la qualité des services fournis par le projet

Analyse de l'accessibilité du projet pour différents secteurs de la population cible

Analyse des facteurs contextuels (conditions économiques, politiques, organisationnelles et environnementales) qui influent sur le projet sur chacun des sites

Compréhension des caractéristiques culturelles des populations concernées et des façons dont celles-ci influent sur la mise en oeuvre du projet et ses effets directs

Recours à la triangulation pour obtenir au moins deux estimations indépendantes des principaux processus et indicateurs d'effets directs (voir ci-dessous).

Les estimations sont toujours plus solides lorsqu'on peut les confirmer au moyen d'au moins deux sources indépendantes. Pour ce faire, on peut :

Se servir d'estimations indépendantes des changements dans les variables d'impact, obtenues au moyen de sondages, observations, groupes de discussion et données secondaires. Comparer les estimations par triangulation. Si les estimations provenant de sources différentes se recoupent, on pourra avoir davantage confiance dans les constatations.

Si les estimations ne se recoupent pas, il faut une stratégie de suivi pour déterminer pourquoi et ajuster les estimations.

Le Chapitre 10 montre comment se servir de la triangulation pour comparer trois estimations indépendantes (sondage, observation et répondants clés) du revenu des ménages. Dans l'un des exemples, les trois estimations sont *convergentes* (consistantes), alors que dans le deuxième les estimations sont *divergentes* (inconsistantes). Dans ce deuxième cas, il est nécessaire de faire un suivi pour déterminer les raisons de l'écart et de

décider comment choisir la ou les valeurs les plus crédibles que l'on utilisera pour l'analyse. Dans l'idéal, le modèle d'évaluation devrait laisser assez de temps et de ressources pour permettre de revenir sur le terrain pour le suivi, soit pendant la phase de supervision des entrevues, soit pendant la phase d'analyse (lorsque les écarts sont découverts, en général). Malheureusement, cela est rarement possible, en particulier avec un budget d'ESR, et on envisagera donc d'autres options :

On peut donner des instructions aux enquêteurs pour qu'ils notent les incohérences entre l'information rapportée et leurs observations directes. Ils devraient indiquer comment ils interprètent les divergences et, si possible, l'estimation qui est la meilleure selon eux. On peut également leur demander de poser quelques questions de contrôle (voir le Chapitre 11).

Les incohérences devraient être repérées pendant la phase de supervision des entrevues et le suivi, au moyen de discussions avec les intervieweurs et, dans certains cas, une seconde rencontre avec un échantillon de répondants.

On devrait définir les règles pour le codage du matériel d'enquête et l'analyse des traitements possibles des incohérences (faudrait-il donner plus de poids à une source de données particulière, par exemple, ou faudrait-il ajuster les estimations quantitatives d'une certaine façon bien définie?).

Pour les variables critiques, il est possible de créer deux indicateurs différents donnant une estimation élevée et une estimation faible (pour le revenu, les effectifs scolaires ou le chômage, par exemple), une pour laquelle l'information provenant de l'enquête n'est pas ajustée et l'autre pour laquelle elle l'est. Les deux estimations seront présentées séparément dans l'analyse.

Faire des ajustements pour tenir compte des différences entre le groupe de projet et le groupe de comparaison

Analyse multi variable

Lorsque de grandes enquêtes par sondages sont effectuées, on procède souvent à une analyse multi variable pour contrôler statistiquement les différences entre les groupes de projet et de comparaison, afin de parvenir à de meilleures estimations de l'impact du projet (voir les Chapitres 10 et 11). Cette analyse consiste à prendre dans chacun des groupes des sujets comparables (individus, ménages, par exemple) de par leur âge, leur éducation et leur revenu, par exemple, et à déterminer s'il existe encore une différence significative entre le groupe de projet et le groupe de comparaison en ce qui concerne la variable d'impact (proportion d'enfants allant à l'école, nombre d'adultes au chômage, par exemple). Si la différence est encore significative après cet appariement, il y a davantage lieu de penser que le projet contribue véritablement à la différence. Si, au contraire, il n'y a plus de différence, alors cela suggère que la scolarisation ou le chômage sont davantage déterminés par les caractéristiques du ménage que par la participation au projet.

Utiliser des méthodes qualitatives pour analyser les différences entre le groupe de projet et le groupe de comparaison

Répondants clés, groupes d'information et observation des participants sont quelques-unes des méthodes que l'on peut utiliser pour identifier les différences qui

pourraient être importantes entre les groupes de projet et de comparaison et pour déterminer comment elles pourraient influencer sur les effets directs du projet et sur l'estimation des impacts du projet (voir le Chapitre 5 pour plus d'information sur la collecte de données auprès de groupes difficiles à atteindre et sur les méthodes rapides pour comparer les groupes de projet et de comparaison).

Doter l'évaluation en personnel de façon économique

Dans cette section, nous traitons des questions relatives aux experts en la matière de l'extérieur (d'un autre pays ou d'une autre région du pays), aux spécialistes du domaine et aux personnes disponibles sur place pour la collecte de données. L'idéal est d'assembler une équipe d'évaluation composée de gens ayant un éventail d'expérience, de compétences et de perspectives différentes. Quand on se heurte aux contraintes associées à l'ESR, en particulier celles associées au financement, il faut parfois faire des compromis au moment de la constitution de l'équipe d'évaluation. Nous traitons ici de chaque catégorie de collaborateurs séparément mais il est important de penser à la composition globale de l'équipe et à l'efficacité de l'équipe d'évaluation au complet pour ce qui est de répondre aux besoins qui existent quand on fait une évaluation.

Faire un usage judicieux des consultants externes

Lorsque l'on fait appel à des consultants externes, c'est généralement (a) parce que l'expertise technique fait défaut au niveau local (au sein de l'organisation ou parmi les chercheurs sur place), (b) pour renforcer la capacité locale, (c) pour gagner du temps, (d) pour garantir indépendance et objectivité, (e) pour être sûr d'avoir des experts crédibles et/ou (f) parce que l'organisme de financement l'exige. Si des consultants externes, lorsqu'ils sont bien utilisés, peuvent considérablement améliorer la qualité des évaluations actuelles et futures, ils coûtent cependant cher et peuvent créer des perturbations. Ils doivent donc être choisis et utilisés avec prudence. Étant données les contraintes qui vont de pair avec une ESR, le but devrait être de limiter le recours à des consultants externes aux domaines où ils sont absolument essentiels. Si la plupart des points ci-dessous concernent l'utilisation dans les pays en développement de consultants en évaluation de renommée nationale ou internationale, les mêmes grands principes s'appliquent aussi aux pays développés. Ainsi, lorsque l'on a un consultant qui prend l'avion de Washington D.C. pour aller faire une évaluation sur la côte Ouest des États-Unis (la distance étant de près de 3 000 milles), par exemple, cela coûte presque aussi cher en temps et en argent que lorsqu'un consultant va en avion d'Angleterre au Niger. Par ailleurs, il y a de nombreuses situations aux États-Unis où l'anglais n'est pas la première langue dans la région du projet et les compétences linguistiques et la sensibilité aux réalités culturelles sont également importantes.

Voici quelques règles générales pour le choix et l'utilisation de consultants :

Veiller à ce que les organismes locaux et le client participent activement à la formulation des exigences pour les consultants externes et au processus de sélection.

Réfléchir attentivement aux mérites d'un consultant international par rapport à un consultant du pays. Si le premier apporte souvent une plus grande expertise technique, le second a généralement une meilleure connaissance de la situation locale (outre le fait, bien sûr, qu'il parle la langue du pays) et il y a des choix à faire. On risque de se mettre le

milieu professionnel local à dos si on ne fait pas du tout appel à des consultants du pays et il pourra être difficile, ensuite, d'obtenir sa coopération.

Si on utilise un consultant international, il faut donner la priorité aux candidats qui ont de l'expérience dans le pays en question et, le cas échéant, connaissent la langue locale.

Pour les évaluations axées sur les opérations, on évitera de choisir des consultants couverts de titres universitaires mais n'ayant guère d'expérience des évaluations de programmes. Le travail sur le terrain nécessite des compétences différentes des projets de recherche universitaire.

Souvent, on n'utilise pas les consultants externes de la façon la plus économique, soit parce qu'ils font beaucoup de choses que le personnel local pourrait faire aussi bien ou mieux, soit parce qu'on les a fait venir au mauvais moment. Voici quelques suggestions pour faire en sorte que les consultants externes soient bien utilisés :

Bien préciser ce que l'on demande au consultant de faire et s'interroger sur la nécessité de s'embarquer dans toutes ces activités.

Même lorsque le budget est serré, essayer de prévoir suffisamment de temps pour que le consultant se familiarise avec l'organisme, le projet et le contexte dans lequel ce dernier est mis en oeuvre. Un consultant qui ne comprend pas le projet, n'a pas passé un peu de temps dans la communauté ou n'a pas tissé de liens avec le personnel, les clients et les autres parties concernées du projet ne servira pas à grand-chose.

Planifier soigneusement à quels stades le consultant devrait intervenir et s'assurer à l'avance qu'il sera disponible au moment voulu. Faire preuve de fermeté avec les consultants qui veulent changer le calendrier, surtout si c'est à la dernière minute, parce que ce serait plus pratique pour eux. Les moments critiques pour l'intervention d'un consultant sont les suivants :

Pendant la phase de définition du champ de l'évaluation où des décisions critiques sont prises quant aux objectifs, à la conception et aux méthodes de collecte des données et où l'on s'entend avec le client sur les solutions possibles pour respecter les contraintes en termes de temps, d'argent et de données.

Lorsque des décisions sont prises quant à la taille et à la composition de l'échantillon

Lorsque l'on examine les résultats de la première phase de la collecte de données

Lorsque l'on prépare le rapport préliminaire de l'évaluation

Lorsque l'on présente les constatations de l'évaluation aux diverses parties concernées.

Prévoir la préparation d'un bref document d'information (étude préparatoire) par le personnel de l'organisme ou les consultants locaux, avant que le consultant externe ne commence à travailler. Ce document devrait résumer l'information importante sur le projet (compilation des documents clés, entre autres), les principaux organismes partenaires et le contexte dans lequel s'inscrit le projet. Il pourra aussi inclure des études

diagnostiques rapides dans quelques collectivités. Ce document devrait être préparé en consultation avec le consultant, par téléphone ou par courriel, par exemple. S'il est bien préparé, un document de ce type peut faire gagner beaucoup de temps aux consultants et amorcer un dialogue sur les grandes questions et priorités entre les clients, l'équipe de recherche locale et les parties concernées avant même que le consultant n'arrive.

Envisager d'utiliser des vidéoconférences ou des conférences téléphoniques pour que le consultant puisse avoir des contacts plus fréquents avec les autres personnes participant à la planification et à la mise en oeuvre de l'évaluation. Cela permettra au consultant d'apporter sa contribution aux stades critiques de l'évaluation sans avoir à être sur place. Il pourra ainsi faire des suggestions suffisamment tôt, pour l'échantillon ou d'autres aspects de l'évaluation, pour qu'il soit possible de suivre ses recommandations et d'apporter des changements. Un autre avantage des vidéoconférences et des conférences téléphoniques est que l'on peut facilement changer leur date si nécessaire. On évitera ainsi les situations extrêmement coûteuses où, par exemple, le consultant se rend en avion d'Europe en Afrique pour participer à la phase de conception de l'évaluation et découvre en arrivant que tout le processus a plusieurs semaines de retard.

Envisager l'embauche de spécialistes dans le domaine

En plus d'experts de certains aspects pertinents de l'évaluation (interviews qualitatives, élaboration de questionnaires, plan d'échantillonnage et analyse de données, par exemple), il est aussi essentiel d'inclure dans l'équipe au moins un membre ayant une bonne connaissance du sujet sur lequel porte l'évaluation (développement agricole, éducation secondaire, ou microcrédit, par exemple). Dans l'idéal, si les ressources le permettent, l'équipe devrait inclure un spécialiste du domaine ayant de l'expérience dans de nombreux pays ainsi que quelqu'un connaissant la situation locale. Le système scolaire ou le système des soins de santé à Chicago ou à Dhaka auront sans doute de nombreuses caractéristiques distinctives (culturelles, organisationnelles et politiques) qu'il est important d'incorporer à l'évaluation.

Nous montrons au Chapitre 14 que la même chose est vraie pour le plan d'échantillonnage et l'analyse statistique. L'efficacité du plan d'échantillonnage et l'application des tests statistiques qui conviennent sont souvent compromises parce que le statisticien n'est pas familiarisé avec les études sur les pratiques dans un domaine particulier (comme les tests de connaissance ou un domaine spécifique de la recherche en psychologie).

Faire preuve de créativité pour les collecteurs de données

Des solutions créatives peuvent parfois être trouvées pour dépenser moins d'argent sur l'embauche de collecteurs de données. Ainsi, pour une évaluation de la santé, il est parfois possible d'embaucher des étudiantes-infirmières et, pour une évaluation agricole, des agents de vulgarisation agricole. Pour de nombreux types d'évaluation, on peut souvent recruter des étudiants comme intervieweurs et enquêteurs. On peut également s'arranger avec l'hôpital d'enseignement local, le ministère de l'agriculture ou un professeur d'université pour embaucher des étudiants ou des employés à un taux de rémunération qui les satisfait tout en étant bien au dessus du tarif normal. Si elles sont séduisantes parce qu'elles permettent parfois de faire des économies ou de développer la capacité locale en

matière d'évaluation, ces solutions ne sont cependant pas sans risque du point de vue de la qualité. Il se peut que les intervieweurs ne prennent pas très au sérieux la tâche qui leur a été confiée. Il arrive aussi qu'il soit difficile, pour des raisons politiques, de ne sélectionner que les intervieweurs les plus prometteurs ou de prendre des mesures contre ceux dont le travail n'est pas satisfaisant. Les coûts peuvent en outre être élevés pour la supervision et la formation et la collecte de données peut prendre plus longtemps. Cela dit, l'expérience montre que ces types de coopération peuvent très bien fonctionner si l'organisme ou la faculté prend la question au sérieux.

Une autre solution créative consiste à embaucher des collecteurs de données de la collectivité. Une école secondaire locale pourra parfois effectuer une évaluation des besoins, ou bien un organisme communautaire pourra faire une étude des conditions de base ou un suivi des progrès du projet. On pourra également utiliser diverses techniques d'auto-vérification. Des personnes ou des familles peuvent par exemple tenir un journal de leurs revenus et dépenses, de leurs heures d'utilisation chaque jour, ou de l'heure, du mode de transport et de la destination de leurs déplacements. On peut donner des appareils photo, des magnétophones ou des caméras à des groupes de la communauté et leur demander de documenter des enjeux comme les problèmes auxquels sont confrontés les jeunes, les besoins de la communauté ou l'état de l'infrastructure au sein de cette dernière. Bien qu'elles puissent affecter la validité des constatations, ces techniques sont utiles pour comprendre le point de vue de la collectivité sur les questions à l'étude.

Collecter les données efficacement

Simplifier les plans pour la collecte des données

La collecte des données est souvent l'un des éléments les plus coûteux, en temps et en argent, de l'évaluation. Tous les efforts pour qu'elle prenne moins de temps ou revienne moins cher sont par conséquent synonymes, dans la grande majorité des cas, avec la simplification des plans pour la collecte des données. Les trois principales approches possibles sont les suivantes :

1. Discuter avec le client de l'information dont on a véritablement besoin pour l'évaluation et éliminer des paramètres ou des discussions ultérieures toute l'information qui n'est pas essentielle pour répondre aux questions clés sur lesquelles est axée l'évaluation.
2. Passer en revue les instruments de collecte de données pour éliminer toute l'information qui n'est pas nécessaire. Ces instruments ont tendance à s'allonger au fil des suggestions faites par les uns et les autres d'ajouter certains éléments qu'il serait « intéressant » d'inclure, même s'ils n'ont pas un rapport direct avec le but de l'évaluation.
3. Simplifier le processus de collecte de données pour que cette dernière coûte moins cher et prenne moins longtemps. On trouvera dans le tableau 3.1, au Chapitre 3, un résumé des différentes stratégies pour réduire les coûts associés à la collecte des données. Parmi ces stratégies :

La simplification de la conception de l'évaluation (l'élimination de la collecte de données de base ou du groupe de comparaison, par exemple)

La précision des besoins du client en matière d'information (comme discuté plus haut)

La recherche de données secondaires fiables (comme discuté plus haut)

La réduction de la taille de l'échantillon (voir le Chapitre 14)
 La réduction des coûts associés à la collecte, la saisie et l'analyse des données (utilisation de questionnaires à remplir soi-même, recours à l'observation directe plutôt qu'à des enquêtes, à des groupes de discussion au lieu de forums communautaires et utilisation de collecteurs de données coûtant moins cher).

Commander des études préparatoires

Il est parfois possible de faire d'importantes économies d'argent et de temps en demandant à un employé de l'organisme ou à un consultant local de faire une étude préparatoire. Celle-ci pourra inclure :

- Une description des différentes composantes du projet sur lequel porte l'évaluation et de la façon dont elles sont organisées
- Des renseignements de base sur l'organisme d'exécution
- Des études préliminaires rapides sur les collectivités du projet et, si possible, une comparaison des collectivités
- De l'information sur les organismes gouvernementaux, les ONG et les autres organismes qui participent au projet ou le connaissent
- Des recommandations quant aux dirigeants communautaires et aux autres répondants clés que le consultant international devrait rencontrer, ainsi que de l'information d'arrière-plan sur eux.

Chercher des données secondaires fiables

On gagnera beaucoup de temps et on économisera beaucoup d'argent si l'on arrive à obtenir des données secondaires fiables. Selon le pays et le sujet, il est parfois possible de trouver des dossiers tenus par des organismes gouvernementaux s'occupant des statistiques ou de la planification, des universités ou d'autres organismes de recherche, des écoles des banques commerciales ou des programmes de crédit, des médias de masse ou d'autres secteurs de la société civile. L'évaluateur devrait en effet se servir de tous les dossiers pertinents comme les données de suivi et les rapports annuels produits par l'organisme d'exécution lui-même. Ces dossiers sont parfois préparés à des fins de planification, de contrôle administratif et financier, d'évaluation des progrès ou encore pour communiquer avec les différents groupes dont l'autorisation, le soutien financier ou l'accord général sont essentiels à la réussite de l'organisation et de ses activités. Certains des points qu'il est important de vérifier lorsque l'on détermine les forces et les faiblesses de divers types de données secondaires sont discutés plus haut dans le présent Chapitre.

Attention : Ne jamais accepter aveuglément les données secondaires sans vérifier leur fiabilité.

Si la plupart des problèmes rencontrés avec les données secondaires tiennent au fait que la période couverte n'est pas la même, que certains secteurs de la population cible sont mal couverts, ou que les données sont mal collectées ou rapportées, il arrive cependant qu'il y ait des faiblesses plus fondamentales dans les données. Dans les cas où la supervision est inadéquate, il se peut qu'un certain nombre d'enquêtes aient été complètement falsifiées par les intervieweurs ou que des sections importantes aient été laissées de côté ou mal enregistrées.

Collecter seulement les données nécessaires

Il est important de s'assurer que seule l'information essentielle est collectée pour éviter d'augmenter les coûts et le temps requis pour la collecte, pour éviter aussi de faire

baisser la qualité de l'information dont on a besoin, les répondants se fatigant quand ils doivent répondre à un grand nombre de questions. Nous recommandons par conséquent que tous les instruments de collecte de données soient soigneusement examinés pour supprimer l'information qui n'est pas pertinente et essentielle pour l'évaluation, et qui ne sera jamais analysée ou utilisée.

De la même façon, le plan d'analyse des données devrait être examiné pour déterminer les types d'analyse des données non regroupées dont on a en fait besoin. Si l'on s'aperçoit que certains des regroupements proposés ne sont pas nécessaires (comparer les impacts du projet sur les participants à plusieurs endroits, par exemple), il sera alors possible de diminuer la taille de l'échantillon.

Trouver des façons simples de collecter les données sur sujets délicats et auprès des populations difficiles à atteindre

Un autre défi pour les évaluateurs, même s'il n'est pas spécifique à l'ESR, est la collecte de données sur des sujets délicats comme la violence familiale, l'utilisation de contraceptifs ou la violence juvénile, ou bien auprès de groupes difficiles à atteindre comme les travailleuses du sexe, les usagers de drogue, les minorités ethniques, les migrants, les sans abri ou encore, dans certaines cultures, les femmes. Diverses méthodes peuvent aider à aborder ces sujets et à atteindre ces groupes. Les contraintes qui vont de pair avec l'ESR – comme le budget, le temps disponible ou les idées politiques – peuvent cependant entraîner des pressions pour que ces sujets délicats ou ces groupes difficiles à atteindre ne soient pas inclus. Il existe au moins trois stratégies pour aborder les sujets délicats :

1. Identifier un large éventail de répondants qui peuvent apporter des perspectives différentes
2. Sélectionner un certain nombre de stratégies adaptées aux cultures concernées pour étudier les sujets sensibles
3. Triangler systématiquement.

Parmi les méthodes adaptées à la réalité culturelle, on pourra choisir dans les méthodes suivantes :

Observation-participation

Observation non participante (observation de personnes ou de groupes en tant que personne de l'extérieur, sans participer à leurs activités)

Groupes de discussion

Études de cas

Répondants-clés

Techniques d'EPRM.

Parmi les groupes difficiles à atteindre, on citera les travailleurs du sexe, les usagers de drogue ou d'alcool, les criminels, les petites entreprises informelles ou non inscrites, les squatteurs et les résidents clandestins, les minorités ethniques ou religieuses, les petits amis ou les pères absents, la main d'œuvre engagée à long terme et les esclaves, les marchands d'eau informels, les filles allant à des écoles pour garçons, les travailleurs migrants et les séropositifs et les sidéens, en particulier ceux qui n'ont pas fait de test de dépistage.

Deux scénarios sont possibles pour l'évaluateur. Dans le premier, on sait que les groupes existent mais leurs membres sont difficiles à localiser et à joindre. Dans le second, les clients et, au moins au début, l'évaluateur ne savent même pas que de tels groupes marginalisés ou « invisibles » existent. Les techniques pour identifier et étudier ces groupes

difficiles à atteindre sont similaires à celles dont on se sert pour aborder les sujets délicats. Elles comprennent :

L'observation- participation. C'est là l'une des façons les plus courantes de se familiariser avec le milieu dans lequel ces groupes opèrent, ou dans lequel on pense qu'ils opèrent, et d'en être accepté. Souvent, les premiers contacts ou les présentations se feront par l'intermédiaire d'amis, de parents, de clients ou, dans certains cas, d'organismes officiels avec lesquels le groupe interagit.

Les répondants-clés. Organiser des interviews avec des gens qui connaissent très bien les groupes cibles et en sont particulièrement proches.

Études par traceurs. On fait appel à des voisins, parents, amis, collègues de travail, etc. pour aider à localiser les gens qui ont déménagé.

Sondage en boule de neige. Avec cette technique, on s'efforce de localiser quelques membres du groupe difficile à atteindre par n'importe quel moyen dont on dispose. On demande ensuite à ces gens d'identifier d'autres membres du groupe. Ainsi, si cela marche, la taille de l'échantillon va augmenter. On utilise souvent cette méthode pour les études de maladies transmissibles sexuellement.

Techniques sociométriques. On demande aux répondants d'identifier les gens vers qui ils se tourneraient pour de l'aide ou des conseils sur des sujets particuliers (conseils en matière de planning familial ou de médecine traditionnelle, ou pour acheter des substances illégales, par exemple). Une carte sociométrique est ensuite dessinée avec des flèches reliant les répondants aux leaders d'opinion, à d'autres répondants ou à des personnes ressources.

Analyser les données efficacement

Chercher des façons de gérer efficacement les données

Avant que les données ne puissent être analysées, elles doivent être saisies électroniquement ou manuellement. Si cela n'est pas fait correctement, la qualité et la fiabilité des données peuvent en pâtir et on peut perdre du temps et gaspiller de l'argent. Qui plus est, si les données ne sont pas gérées comme il faut, des quantités d'information importantes risquent d'être perdues. On trouvera ci-dessous quelques-unes des principales étapes de l'élaboration et de la mise en œuvre d'un plan d'analyse :

Préparer une ébauche de plan d'analyse (voir le tableau 11.4, Chapitre 11).

Pour chaque type d'analyse proposé, ce document doit indiquer les objectifs de l'analyse, l'hypothèse à vérifier, les variables à inclure dans l'analyse et les types d'analyse à effectuer.

Élaborer et tester la liste de codage. S'il y a des questions ouvertes, les réponses doivent être examinées pour définir les catégories qui seront utilisées. Si certaines des données numériques ont été classifiées en catégories (« Plus d'une fois par semaine », « Une fois par semaine », etc.), les réponses devraient être étudiées pour repérer les problèmes ou aberrations éventuels.

Veiller à ce que l'encodage soit fiable. Il faut pour cela s'assurer que la liste de codage est complète et cohérente d'un point de vue logique et aussi

surveiller le processus de codification des données pour veiller à ce que les codeurs ne fassent pas d'erreurs et à ce que leur travail soit uniforme.

Examiner les sondages pour identifier les données manquantes et décider que faire pour ces données (voir Chapitre 14). Dans certains cas, il sera possible de retourner sur le terrain ou de renvoyer les questionnaires aux répondants par la poste, mais la plupart du temps, cela est difficilement faisable. Souvent, les données manquantes ne sont pas aléatoires et il est donc important de se pencher sur ces cas pour éviter les distorsions. Il se peut par exemple que la disposition à répondre à certaines questions varie selon le sexe, l'âge, le statut économique ou le niveau d'éducation des groupes. Il se peut aussi qu'il y ait des différences entre les groupes ethniques ou religieux ou entre les propriétaires des terres et les squatteurs. L'une des premières choses à faire dans l'analyse est de préparer les distributions statistiques des données manquantes pour les variables clés et, lorsque nécessaire, effectuer une analyse préliminaire pour déterminer s'il existe des différences significatives au niveau des données manquantes pour les principaux groupes de population mentionnés ci-dessus.

Entrer les données dans l'ordinateur ou dans le système manuel d'analyse des données.

Nettoyer les données. Il faut pour cela :

Faire une analyse préliminaire des données pour identifier les données manquantes ainsi que les problèmes potentiels comme les observations aberrantes. Ces dernières sont des variables du sondage pour lesquelles certaines notes sont bien au-dessus ou en dessous de la normale pour une donnée particulière. Il suffit de quelques observations aberrantes pour affecter considérablement l'analyse et rendre beaucoup plus difficile l'obtention de résultats statistiquement significatifs (l'écart type étant énormément augmenté). Le processus de nettoyage des données doit par conséquent inclure des règles claires pour le traitement des données aberrantes (voir le Chapitre 11).

Décider comment traiter les données manquantes et comment appliquer les politiques.

Identifier toutes les variables qui pourront avoir besoin d'être ré-encodées.

Fournir la documentation complète de la façon dont les données ont été nettoyées, dont les données manquantes ont été traitées et dont tous les indices ont été créés.

Si les ESR suivent la plupart des procédures normales d'analyse des données, un certain nombre d'affaires peuvent s'imposer lorsque le temps et l'argent sont limités. Lorsque *le temps* est ce qui manque le plus et dans les cas où des ressources supplémentaires peuvent être obtenues pour accélérer le processus, on pourra envisager les approches ci-dessous :

Entrée directe des données d'enquête dans des ordinateurs de poche
Utilisation d'un scanner pour lire les questionnaires

Recours à une université ou à un organisme commercial de recherche pour l'analyse des données
 Embauche de plus de codeurs et analystes de données plus expérimentés

Lorsque *l'argent* est ce qui manque le plus, on pourra choisir parmi les options suivantes :

Limitier les types d'analyse statistique pour réduire le temps passé à l'ordinateur, celui-ci coûtant cher
 Envisager d'acquérir des progiciels populaires pour les statistiques, comme SPSS ou SAS, pour que l'analyse puisse être faite par l'équipe d'évaluation au lieu d'être confiée à des sous-traitants. Inutile de dire qu'il faut pour cela qu'il y ait au sein de l'équipe des gens qui s'y connaissent en statistiques.

Concentrer l'analyse sur les réponses aux grandes questions

Il est toujours sage pour une évaluation de mettre l'accent sur les grandes questions liées à la raison première pour laquelle elle a été entreprise. Cela est particulièrement important pour l'ESR car des choix doivent être faits quant à ce qu'on peut laisser tomber, le temps et le financement étant limités. Lorsqu'on lui rappelle quelles sont les grandes questions et ce dont on a besoin pour y répondre correctement, quiconque planifie une ESR est certain de se concentrer sur ces enjeux et pas sur d'autres. Dans la majorité des cas, les clients et les parties concernées, de même que les évaluateurs, aimeraient collecter d'autres renseignements complémentaires. Néanmoins, lorsque les contraintes qui régissent une ESR entrent en jeu, il faut laisser tomber tout ce qu'« il serait intéressant de savoir » au profit de ce qu'« il est essentiel de savoir » pour répondre aux grandes questions qui sont au cœur de l'évaluation.

Comme expliqué au Chapitre 2 et plus haut dans le présent Chapitre, les grandes questions typiques d'une évaluation ressemblent aux questions suivantes :

- Existe-t-il des preuves que le projet a atteint (ou atteindra) ses objectifs? Y a-t-il des changements mesurables dans les caractéristiques de la population cible en ce qui concerne les impacts escomptés du projet? Quels objectifs ont-ils (et n'ont-ils pas) été atteints? Pourquoi? Est-il raisonnable de postuler que les changements résultent largement du projet plutôt que de facteurs externes (hors du contrôle des personnes qui ont mis le projet en œuvre)?
- Est-ce que le projet visait les bons objectifs? Reposait-il sur un bon diagnostic des causes sous-jacentes du problème (ou des problèmes) à résoudre?
- Quel impact le projet a-t-il eu sur les différents segments de la population cible – y compris les groupes les plus démunis et les plus vulnérables? A-t-il eu des impacts différents sur les hommes et sur les femmes? Y a-t-il des groupes ethniques, religieux ou autres pour lequel le projet n'a pas été bénéfique, ou bien a été néfaste?
- Est-ce que les résultats directs sont durables et est-il probable que les bénéfiques continuent à se faire sentir? Est-ce que les collectivités ou les groupes cibles étaient raisonnablement typiques de populations plus vastes (tous les agriculteurs pauvres ou tous les habitants de taudis

urbains, par exemple) et est-il probable que les mêmes impacts puissent être obtenus si le projet était reproduit à une plus grande échelle?

Quels sont les facteurs contextuels et externes qui ont déterminé l'ampleur de la réussite ou de l'échec du projet?

L'évaluateur qui fait une ESR doit comprendre quelles questions essentielles doivent être étudiées en détail et quelles questions, moins cruciales, peuvent donc être examinées plus superficiellement ou éliminées complètement. Il est en outre essentiel de comprendre quand le client a besoin d'une analyse statistique rigoureuse (et onéreuse) pour pouvoir défendre la légitimité des constatations de l'évaluation devant des membres du congrès ou du parlement ou devant des organismes de financement enclins à critiquer le programme, et quand une analyse et des constatations plus générales suffisent. La réponse à ces questions peut avoir un impact majeur sur le budget de l'évaluation et le temps qu'elle prendra et encore plus sur la composition et la taille de l'échantillon.

Rapporter les constatations avec efficacité et efficacie

Comme mentionné dans la section ci-dessus intitulée « Adapter les plans pour l'évaluation », une évaluation devrait mettre l'accent sur les grandes questions liées à la raison première pour laquelle elle a été entreprise. Cela est particulièrement important pour l'ESR car des choix doivent être faits quant à ce qu'on peut laisser tomber, le temps et le financement étant limités. Il ne faut pas perdre ces grandes questions de vue pendant la planification de l'évaluation, la collecte et l'analyse des données et également au stade de la rédaction du rapport (ou des rapports). Il est souvent tentant de rapporter toutes sortes de « constatations intéressantes », mais les évaluateurs doivent se limiter dans le rapport à répondre aux grandes questions auxquelles les clients et les parties concernées souhaitent voir apporter des réponses.

L'un des moyens les plus efficaces d'augmenter les chances que les constatations de l'évaluation soient utilisées est de s'assurer qu'elles soient d'une utilité pratique et immédiate pour les divers protagonistes. Parmi les facteurs qui influent sur l'utilisation, on citera :

Le moment choisi pour l'évaluation

Reconnaître que l'évaluation n'est que l'une de plusieurs sources d'information et d'influence pour les décideurs et faire en sorte qu'elle vienne compléter les autres sources

Mettre à profit les rapports avec les parties concernées, en prêtant une oreille attentive à leurs besoins, en essayant de comprendre leur perception du contexte politique et en les tenant au courant des progrès de l'évaluation. Il ne devrait y avoir « aucune surprise » lorsque le rapport d'évaluation est présenté (Operations Evaluation Department, 2005; Patton, 1997)

On procédera comme suit pour présenter les constatations de l'évaluation :

Comprendre les parties concernées par l'évaluation et comment elles souhaitent recevoir l'information.

Utiliser une présentation visuelle en complément des rapports écrits ou des présentations orales. Lorsque possible, utiliser des outils de présentation comme PowerPoint, sans devenir esclave de la technologie et en se

préparant à travailler sans ces outils au cas où les questions de logistique deviendraient trop compliquées. Les supports visuels sont particulièrement utiles lorsqu'on présente les constatations dans une langue qui n'est pas la première langue de nombre de gens dans l'auditoire.

Communiquer les résultats de l'évaluation au moyen de présentations orales.

Beaucoup de gens ne sont pas à l'aise avec des rapports écrits ou des diapositives, et il est donc important de parler des constatations.

Planifier un rapport écrit qui sera simple, agréable à l'œil et facile à consulter. Envisager de présenter différentes versions des constatations pour les rendre plus faciles à comprendre et à utiliser pour les différents publics (plus de détails plus loin).

Inclure les médias de masse. Lorsque l'un des buts est d'atteindre et d'influencer un public nombreux (l'opinion publique, tous les parents d'élèves en âge d'être au secondaire, les législateurs, par exemple), la presse peut être une alliée précieuse. Travailler avec les médias prend cependant du temps et une certaine préparation et, si leur participation est importante, cela peut valoir la peine d'engager un consultant qui connaît la question.

Rapport succinct pour les principaux clients

Souvent, les évaluations n'ont pas autant d'impact qu'elles le devraient parce que leurs constatations et leurs recommandations ne sont pas présentées aux principaux clients sous une forme qui leur plaît et qu'ils comprennent. Il n'existe pas *une seule* façon de présenter les constatations d'une évaluation qui soit meilleure que les autres – tout dépend des clients et de la nature de l'évaluation. Un bon point de départ est de demander aux clients quels rapports antérieurs ils ont trouvés utiles et pourquoi.

En règle générale, et en particulier pour les ESR où le temps tend à manquer, il faut que la présentation soit courte et succincte. C'est une bonne idée d'avoir un document papier court que l'on peut distribuer largement. Même si le sommaire exécutif au début d'un long rapport est bien écrit, la taille du document intimide parfois certains clients et intervenants qui risquent de ne même pas l'ouvrir.

Vaughan et Buss (1998) proposent des directives utiles pour déterminer que dire à des décideurs très occupés et comment le leur dire. Ils font remarquer que, si beaucoup d'entre eux ont les capacités intellectuelles nécessaires pour lire et comprendre des analyses complexes, ils n'ont pas le temps de le faire. Ils voudront par conséquent qu'on leur donne un sens des complexités de l'analyse (ils n'aiment pas qu'on leur parle avec condescendance), mais sans se perdre dans les détails. D'autres décideurs qui n'ont pas de connaissances techniques préféreront une présentation plus simple. Il faut donc trouver le juste milieu entre conserver le respect et l'intérêt de ceux qui ont des connaissances techniques sans perdre en route ceux qui n'en ont pas. Il demeure que le temps est précieux pour tout le monde. La présentation doit par conséquent être brève, même si elle n'est par forcément simple. Les règles de Vaughan et Buss pour décider ce qu'il faut dire sont les suivantes :

Analyser les politiques mais pas la politique. Les évaluateurs sont engagés pour fournir des compétences techniques, pas pour donner des conseils sur les stratégies politiques.

Faire simple.

- Communiquer le raisonnement ainsi que les conclusions. Les décideurs veulent souvent savoir comment l'évaluateur est arrivé aux conclusions pour déterminer quel est le poids de ces conclusions.
- Ne pas abuser des chiffres.
- Élucider, ne pas préconiser. S'ils préconisent des politiques particulières, les évaluateurs risquent de perdre la confiance des décideurs.
- Identifier les gagnants et les perdants. Les décideurs se soucient de l'impact des politiques sur les parties intéressées, en particulier à court terme. Par conséquent, si les évaluateurs et les analystes veulent que les décideurs les écoutent, ils doivent identifier les gagnants et les perdants. L'un des points les plus convaincants de l'étude sur les raisons pour lesquelles on devrait mettre un terme au programme de rations de farine de blé au Pakistan, un programme très onéreux mais aux ramifications politiques multiples, était l'analyse montrant qui étaient les perdants potentiels (les distributeurs de farine de blé et les propriétaires des magasins) et comment on pouvait atténuer leurs pertes (Operations Evaluation Department, 2005, chap. 6).
- Ne pas sous-estimer les conséquences involontaires. Les gens accueillent souvent les nouveaux programmes et les nouvelles politiques de façons inattendues, en particulier pour tirer parti des ressources et des débouchés qui se présentent tout à coup. Ces réactions inattendues peuvent détruire un programme qui aurait pu être bon. Elles peuvent aussi contribuer à la réussite du programme. Les responsables des politiques sont sensibles à ces résultats inattendus car ils comprennent qu'ils peuvent coûter très cher sur le plan politique ou économique. Par conséquent, si l'évaluation peut identifier certaines conséquences majeures auxquelles les responsables des décideurs n'avaient pas pensé, cela captera l'attention de l'auditoire et ajoutera à la crédibilité de l'évaluation.

Des rapports pratiques, compréhensibles et utiles pour les autres auditoires

En plus du client et des principaux protagonistes (ministères gouvernementaux concernés et organisme de financement, par exemple), il y a souvent d'autres parties concernées que l'évaluation intéresse pour des raisons différentes. L'évaluation a une incidence directe sur certains groupes, comme les membres de la population cible. D'autres groupes participent à des activités de défense des droits et peuvent souhaiter utiliser les constatations pour appuyer leurs arguments ou critiquer le rapport parce qu'il va à l'encontre de ce qu'ils essaient de faire. D'autres encore sont intéressés par les applications pratiques des constatations. Il arrive souvent que le client ne souhaite pas diffuser trop largement les constatations de l'évaluation, en particulier si celles-ci revêtent une importance critique ou risquent de soulever des questions délicates. Dans ces cas-là, les évaluateurs se retrouveront peut-être confrontés à des préoccupations éthiques et professionnelles délicates. Ils devront décider s'ils sont dans l'obligation morale, voire professionnelle, de communiquer les constatations de l'évaluation à tous les groupes touchés par le projet, même si le client leur a demandé de limiter la distribution de l'information. Ces questions éthiques sont discutées aux Chapitres 6 et 7.

À supposer qu'une solution satisfaisante ait été trouvée à ces questions éthiques, une stratégie de diffusion doit être élaborée pour atteindre les différents groupes, sachant que ceux-ci n'ont pas tous les mêmes centres d'intérêt, les mêmes niveaux de compétence pour lire les rapports d'évaluation et les mêmes préférences en terme de la façon dont ils reçoivent l'information. Dans certains cas, les différents groupes auront aussi besoin que le

rapport leur soit communiqué dans d'autres langues. L'équipe d'évaluation doit décider lesquelles des parties concernées sont suffisamment importantes pour que cela justifie la préparation de versions différentes du rapport (peut-être même de traductions dans d'autres langues) ou l'organisation de présentations et débats séparés.

Ces questions revêtent une importance toute particulière pour les ESR car atteindre des publics différents, en particulier si ceux-ci sont parmi les plus démunis, ceux avec le moins d'éducation et les moins accessibles, coûte considérablement plus cher et prend beaucoup plus de temps. Il y a un risque que, lorsqu'il y a des contraintes financières et temporelles, les résultats de l'évaluation ne soient communiqués qu'aux principaux clients et que les gens dont la vie est la plus touchée (comme les groupes indigènes dont le mode de vie est menacé, les squatteurs urbains qui pourraient être déplacés de force, ou encore les groupes à faible revenu qui ne profiteront pas forcément des nouvelles technologies d'approvisionnement et d'assainissement de l'eau) ne voient jamais l'évaluation ou ne soient jamais consultés au sujet des conclusions et des recommandations.

Pendant cet exercice d'identification du champ de diffusion (Étape 1 de l'ESR), il est important de s'efforcer de parvenir à une entente avec le client qui recevra le rapport d'évaluation et aura l'occasion de partager ses opinions à son sujet. Si cela ne semble pas vraiment intéresser le client de le diffuser plus largement, sans pour autant qu'il n'y voie d'objections, l'évaluateur peut proposer des stratégies abordables pour atteindre un public plus vaste. Si, au contraire, il s'oppose formellement à ce que plus de gens soient consultés ou reçoivent le rapport, l'évaluateur devra alors voir ce qu'il souhaite faire – l'une des options étant de refuser le contrat d'évaluation.

Si le temps et l'argent sont les seuls obstacles à une diffusion plus large, on pourra envisager les options suivantes :

- Obtenir l'aide des médias de masse. Pour cela, il sera souvent nécessaire d'investir un temps considérable dans de bons rapports avec les journalistes de la télévision, de la radio et de la presse écrite. On pourra les inviter à se joindre à des visites sur le terrain ou à des réunions communautaires et leur envoyer des nouvelles intéressantes de temps à autre.
- Obtenir le soutien d'ONG et d'organisations de la société civile. Celles-ci seront souvent partantes pour diffuser l'information mais elles voudront peut-être présenter les constatations selon leur point de vue (qui pourra être fort différent de celui de l'équipe d'évaluation). Il est donc important d'apprendre à connaître les différents organismes avant de leur demander leur aide pour diffuser les rapports.
- Organiser des rencontres avec les organismes des collectivités cibles pour présenter les constatations et recueillir leurs commentaires. Il est important que ces rencontres soient organisées suffisamment tôt dans la préparation du rapport pour que ces opinions et autres renseignements complémentaires puissent être incorporés au rapport final.

Aider les clients à faire bon usage des constatations

Malheureusement, bien trop souvent, l'évaluation est terminée, un rapport formel est rédigé et remis au client et aucune suite n'est donnée à l'évaluation. Si l'on suit les conseils ci-dessus et que l'on inclut le client et les autres protagonistes clés tout au long du processus, on peut espérer que les constatations d'une évaluation seront pertinentes et prises au sérieux. S'il n'y a pas de suivi, cependant, on pourrait avoir l'impression que l'évaluation

n'a servi à rien. Il existe des exemples d'importants organismes donateurs qui, voyant le peu d'usage fait des rapports d'évaluation, ont décidé de tout simplement arrêter de commanditer des évaluations régulières. Ne vaudrait-il pas mieux faire plus d'efforts pour que les évaluations soient conçues spécifiquement pour répondre aux questions clés, qu'elles soient bien faites et, ensuite, qu'il en soit fait un meilleur usage?

L'un des principaux objectifs des ESR est d'aider les intervenants à se concentrer sur ce qui est le plus important et à être aussi efficaces que possible en faisant des évaluations qui apportent quelque chose et qui sont utiles. La dernière étape – l'utilisation – doit être prise en compte lorsque l'on détermine l'efficacité. Si des renseignements ne sont pas utilisés pour orienter des décisions qui aideront à améliorer la qualité du programme et son efficacité, c'est du gaspillage. Autrement dit, ceux qui effectuent des évaluations doivent comprendre que les suites que l'on donne à l'évaluation sont une composante importante du processus.

L'une des façons de procéder consiste à aider le client à préparer un plan d'action qui décrit les mesures que l'on prendra en réponse aux recommandations de l'évaluation puis pour faire un suivi sur la mise en œuvre de ce plan d'action. C'est une démarche logique dans le cas des évaluations formatives puisque les constatations de ce type d'évaluations sont utilisées pour améliorer la mise en œuvre pour la suite d'un projet en cours. Même dans le cas des évaluations récapitulatives, qui ont pour objectif de déterminer dans quelle mesure chacun des effets directs et des impacts souhaités du projet a été obtenu, ou bien dans les situations où le projet que l'on évaluait est terminé, le suivi devrait inclure des mesures pour aider à mettre à profit les enseignements tirés pour orienter les stratégies futures et la conception de projets futurs. Au minimum, les responsables de l'évaluation devraient faire le nécessaire pour assurer que les constatations et les recommandations soient documentées et communiquées correctement aux décideurs actuels et futurs.

Annexe : Principales figures et tableaux extraits des autres chapitres

Tableau 2.2 Les sept modèles quasi-expérimentaux d'évaluation en situation réelle les plus utilisés

Modèle d'évaluation	Début du projet (pré-test) S ₁	Intervention de projet (continue jusqu'à la fin du projet)	Évaluation à mi-parcours ou plusieurs observations pendant la mise en oeuvre S ₂	Fin du projet (post-test) S ₃	Suivi après un projet en place depuis un certain temps (ex post) S ₄	Stade du cycle de projet auquel chacun des modèles d'évaluation peut commencer à être utilisé
LES DEUX MODÈLES D'ÉVALUATION LES PLUS RIGOUREUX						
1. <i>Modèle longitudinal complet, avec observations pre-, à mi parcours, post- et ex-post et avec un groupe de projet et un groupe témoin.</i> C'est le modèle le plus rigoureux d'un point de vue méthodologique mais aussi celui qui coûte le plus cher et prend le plus longtemps. Il permet aussi d'évaluer le processus de mise en oeuvre du projet et d'analyser les tendances. L'affectation aléatoire des sujets est rarement possible. Par conséquent, ce modèle et les suivants utilisent généralement des groupes témoins sélectionnés pour correspondre le mieux possible au groupe de projet.	P ₁	X	P ₂ C ₂	P ₃ C ₃	P ₄ C ₄	Début
2. <i>Modèle avec mesure avant et après le projet et avec un groupe de projet et un groupe témoin.</i> Dans la plupart des cas, c'est le meilleur modèle disponible quand l'évaluation peut commencer au début du projet avec un budget raisonnable et sans contraintes particulières quant à l'accès aux données ou à l'utilisation de groupes témoins.	P ₁ C ₁	X		P ₂ C ₂		Début
CINQ MODÈLES D'ÉVALUATION MOINS RIGOUREUX						
3. <i>Modèle longitudinal tronqué avec mesures avant et après le projet et avec un groupe de projet et un groupe témoin.</i> Observation du groupe de projet et du groupe de contrôle à deux stades ou plus de la mise en oeuvre du projet, mais l'évaluation ne commence pas jusqu'à ce que le projet soit en cours. L'évaluation commence souvent avec l'examen de mi-parcours.		X	P ₁ C ₁	P ₂ C ₂		Mi-parcours
4. <i>Modèle avec mesures avant et après le projet et avec un groupe de projet, combiné avec une analyse post-test du groupe de projet et du groupe témoin.</i> Aucune données de base collectées sur le groupe de contrôle.	P ₁	X		P ₂ C ₁		Début
5. <i>Modèle avec mesure après le projet et avec un groupe de projet et un groupe témoin.</i> Aucune données de base ou sur la situation à mi-parcours ne sont collectées.		X		P ₁ C ₁		Fin
6. <i>Modèle avec mesure avant et après le projet et avec un groupe de projet.</i> Pas de groupe témoin.	P ₁	X		P ₂		Début
7. <i>Modèle avec mesure après le projet et avec un groupe de projet seulement.</i> Aucune données ne sont collectées sur le projet, aucun groupe témoin. C'est le modèle quantitatif le plus faible, mais le plus fréquemment utilisé car il ne coûte pas cher et prend peu de temps		X		P ₁		Fin
Légende	S = Stade du cycle de projet		P = Participants au projet		C = Groupe témoin	
	P ₁ , P ₂ , C ₁ , C ₂ , etc. = Première, deuxième (et dans certains modèles troisième et quatrième) observations du groupe de projet et du groupe témoin dans un modèle d'évaluation donné					
	X = Intervention de projet (un processus plutôt qu'un événement distinct)					

Tableau 10.3 Les forces et les faiblesses des sept modèles quasi-expérimentaux les plus souvent utilisés

<i>Modèle</i>	<i>Avantages</i>	<i>Désavantages</i>
1. <i>Modèle longitudinal complet, avec observations pre-, à mi-parcours, post- et ex-post et avec un groupe de projet et un groupe de comparaison</i>	C'est le modèle le plus rigoureux qui étudie le processus de mise en oeuvre et la durabilité. Peut être nécessaire pour mettre à l'essai un nouveau projet novateur qui, si son impact peut être démontré, sera mis en oeuvre à une beaucoup plus grande échelle	<ul style="list-style-type: none"> • C'est le modèle le plus onéreux et celui qui demande le plus de temps, et donc le plus difficile à mettre en oeuvre
2. <i>Modèle avec mesure avant et après le projet et avec un groupe de projet et un groupe de comparaison</i>	C'est le modèle le plus rigoureux parmi les MQE polyvalents Avec un groupe de contrôle bien choisi, il fournit de bonnes estimations des impacts du projet	<ul style="list-style-type: none"> • Suppose que le groupe de comparaison est raisonnablement similaire au groupe de projet et d'accord pour participer à deux sondages, même s'ils n'ont rien à y gagner
3. <i>Modèle longitudinal tronqué avec mesures avant et après le projet et avec un groupe de projet et un groupe de comparaison</i>	<ul style="list-style-type: none"> • Permet l'observation des processus ainsi que celle des impacts • Modèle raisonnablement rigoureux, en particulier pour les projets où la mise en oeuvre démarre lentement et où l'on ne manque pas grand chose en commençant plus tard 	<ul style="list-style-type: none"> • N'évalue pas la mise en oeuvre du projet • Ne commence pas avant le mi-parcours environ, et n'évalue donc pas à la mise en route du projet et le début de la mise en oeuvre du projet
4. <i>Modèle avec mesures avant et après le projet et avec un groupe de projet, combiné avec une analyse post-test du groupe de projet et du groupe de comparaison</i>	<ul style="list-style-type: none"> • Permet de déterminer si le modèle de projet marche et produit les extrants visés • Permet d'identifier les similarités et les différences entre domaines du projet et domaines similaires • Permet de déterminer dans quelle mesure le projet pourrait être répliqué 	<ul style="list-style-type: none"> • Ne permet pas de déterminer si les différences observées à la fin du projet entre le groupe de projet et le groupe de comparaison sont attribuables au projet ou à des différences préexistantes entre les deux groupes • Ne permet pas de contrôler si des événements de l'histoire locale pourraient influencer sur les effets directs
5. <i>Modèle avec mesure après le projet et avec un groupe de projet et un groupe de comparaison</i>	<ul style="list-style-type: none"> • Permet d'évaluer des projets pour lesquels on a mis en oeuvre des interventions soigneusement testées ou qui se déroulent dans des endroits isolés où d'autres interventions externes n'interfèrent pas 	<ul style="list-style-type: none"> • Ne permet pas d'estimer avec exactitude la magnitude des impacts du projet • Ne permet pas de contrôler le rôle de l'histoire locale • Ne permet pas de déterminer les possibilités de reproduire le projet à plus grande échelle • Ne permet pas d'étudier le processus de mise en oeuvre du projet

Tableau 10.3 Les forces et les faiblesses des sept modèles quasi-expérimentaux les plus souvent utilisés (suite)

6. <i>Modèle avec mesure avant et après le projet et avec un groupe de projet</i>	<ul style="list-style-type: none"> • Donne une estimation approximative des impacts du projet 	<ul style="list-style-type: none"> • Ne permet pas d'estimer avec exactitude la magnitude des impacts du projet • Ne permet pas de contrôler le rôle de l'histoire locale • Ne permet pas la comparaison avec d'autres collectivités • Ne permet pas de contrôler si d'autres variables ont joué un rôle au moyen d'une analyse multivariante
7. <i>Modèle avec mesure après le projet et avec un groupe de projet seulement</i>	<ul style="list-style-type: none"> • Utile pour les études préliminaires lorsque l'on veut avoir une idée générale de l'efficacité du modèle de projet • Donne une première estimation approximative des résultats, en particulier pour des projets petits ou isolés 	<ul style="list-style-type: none"> • Ne permet pas d'obtenir des estimations raisonnablement précises de l'impact du projet • Difficile d'être sûr que les changements observés sont dus au projet et non à d'autres facteurs ou interventions • Ne permet pas de contrôler le rôle d'événements extérieurs • Ne permet pas d'obtenir des données comparatives pour estimer les possibilités de réplique le projet

NOTE : La rigueur de tous ces modèles peut être accrue en les combinant avec le cadre d'évaluation de l'impact et l'analyse des facteurs contextuels présentés au Chapitre 9 et avec certaines des techniques d'ESR discutées plus loin dans ce chapitre. Pour les modèles 1, 2, 3, 4 et 5 qui utilisent des groupes de comparaison, une analyse beaucoup plus rigoureuse peut être faite en utilisant une analyse par régression multiple pour contrôler statistiquement les différences au niveau des caractéristiques du groupe de projet et du groupe de comparaison. Lorsque des données secondaires fiables sont disponibles, ces modèles peuvent en outre être renforcés au moyen de techniques statistiques d'appariement, sur des scores de propension ou des variables instrumentales par exemple.

LISTE DE CONTRÔLE INTÉGRÉE POUR ÉVALUER L'UTILITÉ ET LA VALIDITÉ DES MODÈLES QUALITATIFS, QUANTITATIFS ET MULTI-MÉTHODES¹
A. Preuves à l'appui (et objectivité)²
<i>Les conclusions sont-elles tirées des preuves dont on dispose et la recherche est-elle relativement dépourvue de préjugés?</i>
1. Les méthodes et procédures utilisées pour l'étude sont-elles correctement décrites? Les données pour l'étude sont-elles conservées et disponibles pour une nouvelle analyse?*
2. Les données sont-elles présentées pour appuyer les conclusions?*
3. L'auteur des recherches a-t-il été aussi explicite et conscient de ce qu'il tient pour acquis ainsi que de ses valeurs et biais que possible?*
4. Les méthodes utilisées pour tenir compte de l'effet des biais convenaient-elles?*
5. Des hypothèses concurrentes ou rivales ont-elles été envisagées?
B. Fiabilité et constance
<i>Le processus de l'étude est-il homogène, cohérent et raisonnablement stable dans le temps, d'un chercheur à l'autre et d'une méthode à l'autre? Si des modèles émergents sont utilisés, est-ce que les processus qui font évoluer le modèle sont clairement documentés?</i>
1. Est-ce que les constatations sont fiables, cohérentes et adaptables d'une source de données à l'autre et à travers le temps?
2. Les données ont-elles été collectées dans un éventail large et bien choisi d'endroits, de moments, de répondants, etc.?
3. Tous les travailleurs sur le terrain avaient-ils des protocoles similaires pour la collecte des données?
4. A-t-on procédé à un encodage et à une surveillance de la qualité et les résultats de ces mesures ont-ils été satisfaisants?
5. Ce que rapportent les différents observateurs converge-t-il? Sinon (ce qui est souvent le cas dans les études qualitatives), le problème est-il reconnu et s'efforce-t-on d'y remédier?*** ³
6. A-t-on eu recours à des examens par des pairs ou des collègues?
7. Les conclusions sont-elles sujettes à des « obstacles à la validité du contenu » [voir la section H]? Si oui, les mesures nécessaires ont-elles été prises pour remédier au problème?***
8. Les règles suivies pour confirmer les propositions, hypothèses, etc. ont-elles été communiquées explicitement?
C. Validité interne, crédibilité et authenticité
<i>Les constatations sont-elles crédibles pour les personnes sur qui portait l'étude et pour les lecteurs, et a-t-on un portrait fidèle de ce que l'on étudie?</i>
1. Jusqu'à quel point les descriptions sont-elles informatives sur le contexte et significatives (« denses »)? Fournissent-elles suffisamment d'information pour que l'on ait une description crédible/valide des sujets ou de la situation à l'étude?***
2. Est-ce que ce qui est rapporté sonne vrai, est logique, paraît convaincant? Est-ce représentatif du contexte local?
3. La triangulation entre les sources de données et les méthodes complémentaires produit-elle en général des conclusions convergentes? Si des méthodes qualitatives d'expansion sont utilisées dans les cas où les interprétations ne convergent pas forcément, les différences au niveau des interprétations et des conclusions sont-elles notées et discutées?***
4. Les données présentées sont-elles clairement reliées aux catégories identifiées dans les théories antérieures ou émergentes? Les constatations sont-elles cohérentes entre elles et les concepts sont-ils systématiquement reliés entre eux?
5. Toutes les incertitudes sont-elles identifiées? A-t-on cherché, et trouvé, des preuves négatives? Comment ont-

¹ Sources : Les sections A-E sont adaptées de Miles et Huberman (1994), Chapitre 10 section C. Voir aussi : Guba et Lincoln (1989). Les sections F à I sont adaptées de Shadish, Cook et Campbell (2002). Les éléments en italiques ont été ajoutés par les auteurs du présent document.

² Miles et Huberman considèrent que tous ces éléments sont des indicateurs de l'existence de preuves à l'appui et d'objectivité. Les auteurs de cet ouvrage considèrent cependant que plusieurs catégories sont indicatives de l'existence de preuves à l'appui mais pas de l'objectivité. Ces catégories sont marquées d'un astérisque (*).

³ Les catégories ajoutées ou modifiées par les auteurs du présent ouvrage sont indiquées par deux astérisques (**).

elles été utilisées? Des explications contraires ont-elles été sérieusement envisagées?
6. Les conclusions sont-elles jugées exactes par les chercheurs chargés de la collecte des données?
7. Les constatations sont-elles sujettes à des « obstacles à la validité interne » [voir la section G]? Si oui, a-t-on fait le nécessaire pour remédier au problème?***
8. Les constatations sont-elles sujettes à des « obstacles à la validité statistique » [voir la section F]? Si oui, a-t-on fait le nécessaire pour remédier au problème?***
D. Validité externe, transférabilité et adéquation
<i>Les conclusions sont-elles applicables à d'autres contextes et dans quelle mesure peuvent-elles être généralisées?</i>
1. Les caractéristiques de l'échantillon de gens, situations, processus, etc. sont-elles décrites de façon suffisamment détaillée pour permettre les comparaisons avec d'autres échantillons?
2. Le plan d'échantillonnage permet-il, théoriquement, la généralisation à d'autres populations?
3. Ceux qui font la recherche définissent-ils le champ et les limites des généralisations qu'il est raisonnable de faire à partir de l'étude?
4. Les constatations contiennent-elles suffisamment de « descriptions denses » pour permettre aux lecteurs de déterminer si elles sont transférables?
5. Un éventail de lecteurs rapportent-ils que les constatations coïncident avec leur expérience personnelle?
6. Les constatations confirment-elles, ou recourent-elles, les théories existantes? La théorie transférable est-elle explicitée?
7. Les processus et les constatations sont-ils suffisamment génériques pour être applicables ailleurs?
8. Les séquences narratives ont-elles été conservées? Une théorie générale allant au-delà du cas particulier a-t-elle été élaborée en se servant des séquences?
9. Le rapport suggère-t-il des contextes dans lesquels on pourrait tester plus avant les constatations et obtenir des résultats intéressants?
10. Les constatations ont-elles été reproduites dans d'autres études pour tester leur solidité? Sinon, pourrait-on s'organiser rapidement pour les reproduire?
11. Les constatations sont-elles sujettes à des obstacles à la validité externe [voir section I]? Si oui, a-t-on pris les mesures nécessaires pour remédier au problème?
E. Utilisation/Application/Accent mis sur l'action
<i>Dans quelle mesure les constatations ont-elles été utiles pour les clients, les chercheurs et les collectivités étudiées?</i>
1. Les constatations sont-elles accessibles pour les utilisateurs potentiels – intellectuellement et matériellement?
2. Des prévisions ont-elles été faites dans l'étude et, si oui, dans quelle mesure sont-elles exactes?
3. Les constatations fournissent-elles des conseils pour des interventions futures?
4. Les constatations ont-elles un effet catalyseur menant à des mesures spécifiques?
5. Les mesures prises ont-elles réellement aidé à résoudre des problèmes locaux?
6. Les utilisateurs des constatations se sont-ils sentis plus autonomes ou plus maîtres de leur vie? Ont-ils acquis de nouvelles compétences?
7. Des préoccupations concernant les valeurs ou la morale ont-elles été mentionnées explicitement dans le rapport? Sinon, y a-t-il de telles préoccupations que les responsables de la recherche ont négligées?
F. Obstacles à la validité statistique des conclusions⁴
<i>Quelles sont les raisons pour lesquelles les inférences quant à la covariation entre deux variables peuvent être inexactes ?</i>
1. <i>Échantillon trop petit pour détecter les effets du programme (efficacité statistique faible)⁵.</i> L'échantillon est trop petit pour qu'il soit possible de détecter des différences statistiquement significatives entre les groupes, même quand différences il y a.
2. <i>Dérogations aux hypothèses des tests statistiques.</i> Pour de nombreux tests statistiques, il faut que les observations soient indépendantes les unes des autres. Si l'on déroge à ce principe, en étudiant des enfants dans la même classe ou des patients de la même clinique, par exemple, autrement dit des personnes qui ont plus de similarités entre eux que l'ensemble de la population, on peut augmenter le risque d'erreurs de Type 1 (voir le chapitre 16) et conclure à tort que le projet a eu un effet.
3. <i>Recherche délibérée de résultats statistiquement significatifs.</i> Un certain pourcentage des tests statistiques auront des résultats « significatifs » par hasard (1 résultat sur 20 si l'on prend pour niveau de signification 0,05). Quand on « part à la pêche » dans de nombreux tableaux statistiques, on trouve toujours

⁴ Les sections F, G, H et I de cette liste sont adaptées de Shadish, Cook et Campbell (2002), des tableaux 2.2, 2.4, 3.1 et 3.2 respectivement. Les éléments marqués de ** ont été ajoutés par les auteurs du présent ouvrage pour leur pertinence toute particulière pour les ESR.

⁵ La terminologie technique a parfois été simplifiée. Les titres d'origine sont indiqués entre parenthèses.

quelques-uns de ces résultats fallacieux.
4. <i>Mesures non fiables du changement dans les indicateurs d'effets directs</i> . Les mesures non fiables, par exemple, des taux de changement au niveau du revenu, de l'alphabétisation et de la mortalité infantile, diminuent toujours la probabilité de trouver un effet significatif.
5. <i>Restriction de la portée</i> . Si la comparaison ne porte que sur des groupes similaires, les tests deviennent moins révélateurs et la probabilité de trouver un effet significatif diminue elle aussi.
6. <i>Mise en oeuvre du traitement non fiable</i> . Si le traitement n'est pas administré de façon identique à tous les sujets, la probabilité de trouver un résultat significatif diminue.
7. <i>Événements extérieurs qui influent sur les effets directs (Variances perturbatrices dans les conditions expérimentales)</i> . Il arrive que les sujets soient distraits par des événements ou des pressions extérieurs (panne de courant, violence au sein de la communauté, campagnes électorales) qui influent sur leur comportement et sur les résultats directs du programme.
8. <i>Diversité de la population (Hétérogénéité des unités)</i> . Si les sujets ont des caractéristiques très différentes, cela peut accroître la variance des résultats et compliquer la détection d'effets significatifs.
9. <i>Estimation inexacte de l'ampleur de l'effet</i> . Quelques observations aberrantes (valeurs extrêmes) diminuent parfois l'ampleur de l'effet de façon significative (voir le Chapitre 16), réduisant ainsi les chances de trouver des différences significatives.
10. <i>Extrapolation à partir d'une base de données tronquée ou incomplète**</i> . Si l'échantillon ne couvre qu'une partie de la population (les familles les plus pauvres par exemple, ou seulement les gens qui travaillent dans le secteur structuré), cela peut affecter les conclusions de l'analyse et biaiser les généralisations à l'ensemble de la population.
11. <i>Des échantillons du groupe de projet et du groupe de comparaison qui ne couvrent pas les mêmes populations**</i> . Souvent, on ne choisit pas l'échantillon du groupe de comparaison au sein d'exactly la même population que l'échantillon du groupe de projet. Dans ces cas-là, les différences au niveau des effets directs peuvent résulter des différences entre les caractéristiques des deux échantillons, et non des effets du projet.
12. <i>Information qui n'est pas collectée auprès des bonnes personnes ou en omettant d'interviewer certaines catégories de répondants**</i> . Il arrive que l'information ne soit pas collectée dans certains secteurs de la société cible, ou à leur sujet (les hommes mais pas les femmes, les enseignants mais pas les élèves), les estimations pour la population cible pouvant alors être biaisées.
G. Obstacles à la validité interne
<i>Quelles sont les raisons pour lesquelles les inférences qu'il y a des rapports de cause à effet entre deux variables sont parfois inexactes ?</i>
1. <i>Séquence peu claire – pas clair si l'activité de projet a eu lieu avant l'effet présumé (Précédence temporelle ambiguë)</i> . Une cause doit précéder son effet. Il est souvent difficile de savoir, cependant, dans quel ordre les événements se sont succédés dans un projet. De nombreux projets (les programmes de développement urbain, par exemple), au lieu d'avoir une date de départ précise, mettent plusieurs mois, voire plusieurs années, à se mettre en route.
2. <i>Sélection</i> . Les participants au projet sont souvent différents de ceux du groupe de comparaison, soit parce qu'ils sont auto-sélectionnés, soit parce qu'on a sélectionné pour le projet des gens ayant des caractéristiques particulières (les fermiers les plus pauvres ou les collectivités les mieux organisées).
3. <i>Événements historiques</i> . La participation à un projet déclenche parfois des expériences sans rapport avec le traitement du projet qui peuvent faire une différence entre le groupe de projet et le groupe de comparaison. Par exemple, les entrepreneurs dont on sait qu'ils ont reçu un prêt sont peut-être plus à risque d'être volés ou poursuivis par des politiciens qui veulent qu'ils fassent une donation, par exemple, ou encore les filles inscrites au secondaire risqueront davantage de tomber enceintes.
4. <i>Maturation</i> . Lorsque les gens prennent de l'âge, on constate des changements dans leur comportement, leur savoir et leur exposition à de nouvelles expériences. Il est souvent difficile de séparer les changements attribuables à la maturation de ceux dus au projet.
5. <i>Régression vers la moyenne</i> . Si les sujets sont sélectionnés pour leurs scores extrêmes (poids, développement physique), ils ont naturellement tendance à se rapprocher de la moyenne avec le temps – ce qui a pour effet de diminuer ou distordre les effets du programme.
6. <i>Attrition</i> . Même quand les personnes participant à l'origine au programme avaient des caractéristiques similaires à celles de l'ensemble de la population, des abandons sélectifs au fil du temps peuvent changer les caractéristiques de la population du projet (il se peut par exemple que les personnes plus pauvres ou celles ayant le moins d'éducation abandonnent).
7. <i>Effet des tests</i> . Le fait d'être interviewé ou testé peut modifier le comportement ou les réponses. Lorsque l'on pose des questions aux gens sur leurs dépenses, par exemple, cela peut les encourager à réduire les

dépenses que la société réproouve (cigarettes ou alcool) et à dépenser plus pour des choses plus acceptables.
8. <i>Évolution de la façon dont les chercheurs décrivent ou interprètent les données au fur et à mesure qu'ils ont plus d'expérience (biais attribuable aux instruments)</i> Lorsqu'ils acquièrent plus d'expérience, les intervieweurs changent parfois la façon dont ils interprètent les échelles d'évaluation, les grilles de contrôle, etc.
9. <i>Distorsion lorsque les répondants font appel à leurs souvenirs (biais de rappel)**</i> . Il arrive que les répondants, délibérément ou intentionnellement, déforment leurs souvenirs d'événements antérieurs. Des politiciens des partis d'opposition pourront exagérer les problèmes dans la collectivité ou bien les aînés de cette collectivité pourront idéaliser le passé.
10. <i>Utilisation de modèles moins rigoureux du fait de contraintes budgétaires ou temporelles**</i> . Les obstacles à la validité interne sont multiples et fréquents pour les ESR, du fait que l'on a besoin de réduire les frais et de gagner du temps.
H. Obstacles à la validité conceptuelle
<i>Quelles sont les raisons pour lesquelles les inférences quant aux concepts qui caractérisent la façon dont l'étude est menée sont peut-être inexactes ?</i>
1. <i>Mauvaise explication du concept</i> . Les concepts (les effets) à l'étude sont définis de façon trop générale pas assez clairement, ou encore sont ambigus, ce qui empêche de les mesurer précisément (sont par exemple ambigus des termes comme chômeur, agressif, milieu de travail hostile, discrimination sexuelle)
2. <i>Indicateurs qui ne mesurent pas correctement le concept (concept prêtant à confusion)</i> . La définition opérationnelle ne rend pas toujours compte comme il faut du concept désiré. En définissant les chômeurs comme ceux qui se sont inscrits auprès d'un centre d'emploi, par exemple, on ignore tous les gens qui sont au chômage mais n'utilisent pas ces centres. De la même façon, quand on définit la violence familiale comme les cas rapportés à la police, on sous-estime considérablement le nombre réel d'incidents.
3. <i>Utilisation d'un seul indicateur pour mesurer un concept complexe (Biais attribuable à une opération unique)</i> . Un seul indicateur (opération) d'un concept complexe (comme la pauvreté, le bien-être, la violence familiale) biaisera généralement les conclusions.
4. <i>Utilisation d'une seule méthode pour mesurer un concept (Biais attribuable à une seule méthode)</i> . Si on se sert d'une seule méthode pour mesurer un concept, cela produira une mesure étroite et souvent biaisée (si l'on observe les collectivités lors de réunions structurées, par exemple, on n'obtiendra pas les mêmes résultats que si on les observait dans le cadre d'activités sociales ou de projets de travail communautaire).
5. <i>Étude d'un seul niveau du traitement (Confusion entre le contenu et les niveaux de contenu)</i> . Souvent, une intervention n'est administrée qu'à un seul niveau d'intensité et celui-ci est souvent assez bas (prêts aux petites entreprises seulement, court programme de formation pour les chefs de file communautaires seulement, don d'un petit nombre d'ordinateurs à une école seulement), et les résultats sont utilisés pour tirer des conclusions générales quant à l'efficacité (ou inefficacité) du traitement. Cette approche est trompeuse, un niveau de traitement plus élevé pouvant avoir un effet plus significatif.
6. <i>Différence entre les façons dont les participants au programme et les membres du groupe de comparaison répondent à certaines questions (Structure factorielle tenant compte du traitement)</i> . Les participants au programme peuvent répondre de façon plus nuancée aux questions. Ils pourront par exemple faire la distinction entre différents types et différentes intensités de violence familiale ou de préjugé racial, le groupe de comparaison ayant pour sa part des réponses plus générales, moins nuancées.
7. <i>Évaluations de leur situation et d'eux-mêmes par les participants qui diffèrent de celles du groupe de comparaison (Changements réactifs dans les auto-évaluations)</i> . Les personnes sélectionnées pour les programmes s'évaluent parfois différemment de celles sélectionnées avant que le programme ne commence. Il est possible qu'elles veuillent exagérer le besoin qu'elles ont du programme (plus pauvres, plus malades) ou bien se présenter comme plus méritantes, si cela fait partie des critères de sélection.
8. <i>Réactivité à la situation expérimentale</i> . Lorsque les participants au projet essaient d'interpréter la situation du projet, cela influe sur leur comportement. S'ils pensent que le programme est offert par un organisme religieux ou un groupe politique, leur réaction risque d'être différente.
9. <i>Attentes de l'expérimentateur</i> . Il arrive que les expérimentateurs aient eux aussi des attentes (quant à la façon dont les hommes et les femmes, ou bien des groupes socio-économiques différents, réagiront au programme) et cela peut avoir une incidence sur la façon dont ils interagissent avec les différents groupes.
10. <i>Nouveauté et effets perturbateurs</i> . Les programmes inédits génèrent parfois une certaine excitation et ont donc un grand effet. Si un programme similaire est introduit à une date ultérieure, l'effet risque d'être moindre, la nouveauté n'étant plus là.
11. <i>Effets de compensation et rivalité</i> . Il arrive que les programmes créent une dynamique qui influe de diverses façons sur les effets directs. Il peut y avoir des pressions (politiques ou autres) pour que les non-participants aient également quelque chose à gagner; des rivalités peuvent naître avec les groupes de comparaison qui essaieront alors de montrer ce qu'ils peuvent faire sans aide; ou encore ceux qui ne sont pas

touchés par le projet ou qui relèvent d'un projet moins attrayant peuvent se démoraliser.
12. <i>Utilisation d'indicateurs et de concepts élaborés dans d'autres pays, sans essais préliminaires dans le contexte local**.</i> Pour de nombreuses évaluations, des théories et des concepts sont importés d'autres pays et ne rendent pas compte comme il faut du projet dans la situation locale. Ainsi, pour beaucoup d'évaluations des impacts du microcrédit sur le renforcement de l'autonomie des femmes dans des pays comme le Bangladesh et l'Inde, on s'est servi de définitions du renforcement de l'autonomie qui ne conviennent pas forcément pour des Bangladaises.
I. Obstacles à la validité externe
<i>Quelles sont les raisons pour lesquelles les inférences quant à la façon dont les résultats de l'étude tiendraient le coup en dépit des variations dans les gens, les contextes, les projets et les effets ne sont pas forcément exactes ?</i>
1. <i>Échantillon qui ne couvre pas l'ensemble de la population à laquelle on s'intéresse.</i> Il arrive que les sujets soient tous du même sexe ou des mêmes groupes ethniques ou économiques, ou bien qu'ils aient les mêmes caractéristiques en termes de personnalité (déprimés, assurés, par exemple). Il peut par conséquent être difficile de généraliser les constatations de l'étude à l'ensemble de la population.
2. <i>Effets directs du programme influencés par la situation (Interaction entre les rapports de cause à effet et les variations dans le traitement).</i> Les projets peuvent être mis en oeuvre dans des situations différentes qui peuvent influencer sur les effets directs. Si les pressions pour réduire la taille des classes forcent les écoles à construire d'autres salles de classe temporaires et de qualité inférieure, les effets risquent de ne pas du tout être les mêmes que si l'on avait des classes plus petites dans des locaux convenables.
3. <i>Mesures différentes des effets qui donnent des évaluations différentes (Interaction entre les rapports de cause à effet et les effets directs).</i> Des mesures différentes des effets directs peuvent mener à des conclusions différentes quant à l'efficacité du projet. Les programmes de microcrédit pour les femmes peuvent entraîner une augmentation du revenu du foyer et des dépenses pour l'éducation des enfants, sans pour autant renforcer les pouvoirs des femmes sur le plan politique.
4. <i>Variation des effets directs du programme selon la situation (Interaction entre les rapports de cause à effet et la situation).</i> La réussite du programme ne sera pas forcément la même en situation rurale et en situation urbaine ou dans des collectivités de types différents. Il ne sera donc pas toujours bon de généraliser les constatations à des contextes différents.
5. <i>Fonctionnement différent des programmes selon la situation (facteurs médiateurs associés à la situation).</i> Les programmes pourront fonctionner différemment et avoir des effets intermédiaires et ultimes différentes situations. La mise en oeuvre d'écoles gérées par la collectivité pourra se faire différemment et avoir des effets différents selon que ce sont des organisations religieuses, gouvernementales ou non gouvernementales qui s'en occupent.
6. <i>Attitude des décideurs et des politiciens à l'égard du programme **.</i> Des programmes identiques fonctionneront différemment et auront des effets différents dans les situations où ils bénéficient du soutien actif des décideurs ou des politiciens et dans celles où ils sont accueillis avec hostilité ou indifférence. Lorsque le parti au pouvoir ou le directeur de l'organisme change, il est fréquent de voir l'appui reçu par le programme disparaître ou augmenter.
7. <i>Cycles saisonniers et autres**.</i> De nombreux projets fonctionnent différemment selon la saison, le stade du cycle économique et les conditions internes pour les principales importations et exportations. Il faut tenir compte de ces cycles quand on essaye de généraliser les constatations des programmes expérimentaux.