



ROYAUME DE BELGIQUE

Service public fédéral

Affaires étrangères,
Commerce extérieur et

Coopération au Développement

Evaluer l'impact, la quête du Graal ?

Evaluation ex post de l'impact de quatre
projets de la coopération gouvernementale

Rapport de synthèse



inputs outputs outcomes impact

**Service public fédéral Affaires étrangères, Commerce extérieur et
Coopération au Développement**

Service de l'Évaluation spéciale de la Coopération internationale

Mesurer l'impact, la quête du Graal ?

Evaluation ex post de l'impact de quatre projets de coopération gouvernementale



Rapport de synthèse

Octobre 2013

La présente évaluation a été réalisée par ADE et le CRED avec l'appui d'un comité d'accompagnement. Le service de l'Évaluation spéciale a assuré la conformité de l'évaluation aux termes de référence.

Les opinions exprimées dans ce document représentent les points de vue des auteurs et ne reflètent pas nécessairement celles du SPF Affaires étrangères, Commerce extérieur et Coopération au Développement.

© SPF Affaires étrangères, Commerce extérieur et Coopération au Développement

Janvier 2014

Le présent rapport doit être cité comme suit:

Service de l'Évaluation spéciale (2013), *Evaluer l'impact, la quête du Graal ? – Evaluation ex post de l'impact de quatre projets de coopération gouvernementale – Rapport de synthèse*, SPF Affaires étrangères, Commerce extérieur et Coopération au développement, Bruxelles.

Production graphique : Service Communication SPF

Impression : Service Imprimerie SPF

Evaluation N° S4/2011/04

Dépôt légal : 0218/2014/002

Ce document est disponible en format PDF en français, néerlandais et anglais sur le CD-ROM en annexe de la synthèse de l'évaluation, auprès du Service de l'Évaluation spéciale ou sur le site

http://diplomatie.belgium.be/fr/politique/cooperation_au_developpement/evaluation

Nous tenons à remercier ici toutes les personnes et organisations qui ont contribué à cette étude. Nous remercions en particulier les membres du comité d'accompagnement et les personnes interrogées dans le cadre des 4 évaluations ex post d'impact réalisées.

Table des matières

TABLE DES MATIÈRES	5
LISTE DES FIGURES ET DES TABLEAUX	7
ACRONYMES ET ABRÉVIATIONS.....	9
RÉSUMÉ EXÉCUTIF.....	11
1. INTRODUCTION	19
2. MÉTHODOLOGIE.....	21
2.1 L'impact, une définition à clarifier	21
2.2 Un processus évaluatif, deux méthodes.....	23
2.3 Méthodes quantitatives : objectifs et défis	25
2.4 L'organisation pratique de l'évaluation	38
3. PRINCIPALES CONCLUSIONS SUR L'ATTEINTE DES OUTCOMES ET DE L'IMPACT DES 3 PROJETS	41
3.1 Evaluation d'un projet d'appui à l'enseignement technique et professionnel en RDC (AETP2)	41
3.2 Evaluation d'un projet d'amélioration et de renforcement des points d'eau au Sénégal (PARPEBA)	42
3.3 Evaluation d'impact d'un projet de petite et moyenne hydraulique au Maroc (PMH)	43
4. ENSEIGNEMENTS DE L'UTILISATION DE MÉTHODES MIXTES POUR L'ÉVALUATION D'OUTCOMES ET D'IMPACT	49
4.1 Enseignements sur le processus d'évaluation	49
4.1.1 Un dispositif d'évaluation quantitative enrichit l'analyse qualitative ..	51
4.1.2 Une bonne analyse qualitative est nécessaire à la mise en œuvre rigoureuse des méthodes quantitatives	53
4.1.3 L'analyse qualitative indispensable pour comprendre les effets d'un projet (« outcomes »/impact) et pour apprécier l'impact.....	56
4.1.4 Un dispositif qui permet de considérer l'ensemble des acteurs.....	57
4.1.5 Faisabilité et pertinence d'une évaluation quantitative ex post.....	57
4.2 Enseignements sur la conception, la mise en œuvre et le suivi-évaluation des projets	65

4.2.1	Des insuffisances dans la conception/formulation des interventions qui réduisent leur impact potentiel.....	65
4.2.2	Une mise en œuvre qui privilégie le décaissement et les réalisations physiques.....	67
4.2.3	Un système de suivi-évaluation insuffisant.....	67
5.	RECOMMANDATIONS.....	69
5.1	Définir une stratégie globale d'évaluation où l'évaluation d'impact a une place.....	69
5.1.1	Clarifier les définitions de l'impact, de l'« outcome » et de l'évaluation d'impact.....	71
5.1.2	Faire connaître l'évaluation d'impact.....	71
5.1.3	Prévoir un programme annuel d'évaluations d'impact en s'assurant de leur faisabilité et de leur valeur ajoutée.....	72
5.1.4	Garantir la qualité des évaluations.....	73
5.2	Réaliser des évaluations d'impact rigoureuses et de qualité.....	74
5.2.1	Concevoir l'évaluation au moment de la conception du projet.....	74
5.2.2	Etapes successives et éléments importants pour mener une évaluation ex post d'impact rigoureuse.....	75
5.3	« Management towards outcomes and impact » pour des projets plus réalistes et plus d'effets tangibles.....	78
5.3.1	Intégrer l'évaluation d'impact dès la conception et renforcer le processus de validation des projets pour ne pas démarrer des projets irréalistes.....	79
5.3.2	Mettre en pratique une véritable gestion axée sur les résultats (« outputs mais aussi <i>outcomes</i> et impact).....	80
6.	BIBLIOGRAPHIE.....	81
	ANNEXES.....	83
	ANNEXE 1 RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PROJET AETP2 EN RDC	
	ANNEXE 2 RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PARPEBA AU SÉNÉGAL	
	ANNEXE 3 RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PROJET PMH AU MAROC	

Liste des figures et des tableaux

LISTE DES FIGURES

FIGURE 1 : DEUX CONCEPTIONS DE L'ÉVALUATION D'IMPACT	22
FIGURE 2 : DIFFÉRENTES DÉFINITIONS DES CONCEPTS DU CADRE LOGIQUE.....	22
FIGURE 3 : TECHNIQUE DE LA DOUBLE DIFFÉRENCE SCHÉMATISÉE.....	35
FIGURE 4 : ETAPES CLEFS DU PROCESSUS D'ÉVALUATION PAR DES MÉTHODES MIXTES	38

LISTE DES TABLEAUX

TABLEAU 1 : PRÉSENTATION SYNTHÉTIQUE DES 3 RAPPORTS D'ÉVALUATION.....	45
TABLEAU 2 : CONDITIONS POUR LA RÉALISATION D'ÉVALUATIONS QUANTITATIVES DE CONCEPTION QUASI-EXPÉRIMENTALE, EX ANTE OU EX POST À L'INTERVENTION.	63

Acronymes et abréviations

3ie	International initiative for impact evaluation
ADE	Aide à la décision économique
AfD	Agence française pour le développement
AGR	Action génératrice de revenu
ASUFOR	Association des usagers des forages
AUEA	Association des usagers des eaux agricoles
AETP2	Projet d'appui à l'enseignement technique et professionnel en RDC
CAD	Direction de la coopération au développement (OCDE)
CL	Cadre logique
CRED	Centre de recherche en économie du développement
CTB	Coopération technique belge
DGD	Direction générale Coopération au développement et Aide humanitaire
ETP	Enseignement technique et professionnel
EUDN	European Development Network
IEG	Independent Evaluation Group (World Bank)
J-Pal	Abdoul Latif Jamel Poverty action lab
LRSP	Projet de laboratoire de référence en santé publique
MIT	Massachusetts Institute of Technology
MM	Méthodes mixtes
OCDE	Organisation de coopération et de développement économiques
OMD	Objectif du millénaire
PARPEBA	Projet d'amélioration et de renforcement des points d'eau dans le bassin arachidier au Sénégal
PMH	Projet de petites et moyennes hydrauliques dans la région de Tiznit au Maroc
RCT	Randomized Control Trial
RDC	République démocratique du Congo
SE	Suivi et évaluation
SES	Service de l'évaluation spéciale

Résumé exécutif

Introduction

L'évaluation ex post de l'impact de quatre projets de coopération gouvernementale a été commanditée par le Service de l'Évaluation spéciale (SES) de la coopération internationale du Service public fédéral belge des affaires étrangères, commerce extérieur et coopération au développement. Elle a été menée par des experts du bureau d'étude ADE, spécialisés en évaluation et des chercheurs du centre de recherche en économie du développement (CRED de l'université de Namur), spécialisés dans l'application de méthodes quantitatives d'évaluation. L'évaluation portait sur quatre projets dans quatre pays et secteurs différents.

Il s'agit des premières évaluations d'impact strictu sensu commanditées par le SES dont le processus d'évaluation combine des méthodes qualitatives et des méthodes quantitatives pour apprécier les *outcomes* et l'impact de projets.

L'objectif de cette évaluation est à la fois sommatif et formatif. D'une part, le présent rapport résume les enseignements sur l'atteinte des *outcomes* et de l'impact des projets évalués. D'autre part, il tire les leçons sur l'efficacité des méthodes utilisées. Il amorce également une réflexion sur la place de l'évaluation d'impact dans la coopération bilatérale belge et propose des pistes méthodologiques pour les réaliser de façon plus rigoureuse.

Nécessité de préciser les concepts

Un des premiers défis de l'évaluation a été de revisiter la notion d'impact. Selon l'OCDE, la dimension temporelle caractérise l'impact défini comme les « effets à long terme positifs ou négatifs, induits par une action de développement, directement ou non, intentionnellement ou non ». Dans l'approche basée sur la logique d'intervention, l'impact est défini comme étant le dernier élément d'une chaîne causale partant des activités et aboutissant à l'impact. Dans la communauté pratiquant l'*impact evaluation*, comme par exemple la Banque Mondiale et 3ie (*International Initiative for Impact Evaluation*), l'impact est défini comme les effets sur les bénéficiaires qui sont attribuables à l'intervention. Cette dernière définition de l'impact fait donc référence aux *outcomes* tels que présentés dans le cadre logique d'un projet.

Le rapport propose de définir **l'impact** comme étant **l'ensemble des effets induits du projet à un niveau global** (sans référence à leur temporalité) ; et **les « outcomes »**, comme **les effets sur les bénéficiaires directs**.

Cette distinction entre l'impact et les *outcomes* est importante, en particulier lorsque l'évaluation combine des méthodes d'analyse qualitatives et quantitatives. En effet, les méthodes quantitatives mesurent les effets sur les bénéficiaires (« outcomes ») tout en s'efforçant de démontrer les liens de cause à effet entre le projet et les effets mesurés. Pour mesurer et démontrer les effets attribuables au projet, les méthodes quantitatives ont besoin de données fiables en quantité suffisante reflétant la situation préalable à l'intervention (« baseline ») et la situation post-intervention pour des indicateurs d'*outcomes* pertinents. De plus, il est nécessaire de collecter ces données d'une part, auprès d'un échantillon de la population soumise au projet (appelé groupe traité ou groupe cible) et d'autre part, auprès d'un échantillon d'une population qui lui est similaire selon un nombre important de caractéristiques préalables à l'intervention, mais qui ne bénéficie pas de l'intervention (appelé

contrefactuel, groupe de contrôle ou groupe témoin).¹ Les méthodes quantitatives ne se prononcent généralement pas sur l'impact, défini comme les effets induits du projet à un niveau global. Elles se limitent à l'appréciation des « *outcomes* », dont les constats peuvent parfois être généralisés ou extrapolés afin d'en déduire une mesure d'impact.

Les méthodes qualitatives, quant à elles, portent un jugement argumenté sur l'atteinte des *outcomes* et de l'impact des projets. Elles analysent, sur base d'une étude documentaire, d'entretiens à différents niveaux et d'observations de terrain, si les conditions pour atteindre les *outcomes* et l'impact sont remplies aux différents stades de la logique d'intervention résumée dans le cadre logique.

Un processus d'apprentissage

Cette étude se distingue par son côté **pilote et itératif**.

Pilote, car les quatre évaluations réalisées ne sont pas des évaluations classiques de projet, leur objectif central est d'apprécier les *outcomes* et l'impact ex post des projets par une approche méthodologique mixte. Venant de deux mondes très différents et utilisant des techniques d'analyses propres à leur discipline, les experts du bureau de consultance et ceux du centre de recherche académique ont réalisé un travail important de mise en commun et en cohérence de leurs constats. Ces échanges entre les experts ainsi qu'avec le SES et le comité d'accompagnement ont favorisé un apprentissage mutuel. Ils ont permis d'aboutir in fine à un rapport d'évaluation par projet de qualité, ainsi qu'à la rédaction de ce rapport de synthèse. Itératif, car les quatre projets ont été évalués selon cette approche méthodologique dans des pays et secteurs différents afin d'en tirer un maximum d'enseignements tant sur le plan sommatif que formatif. Notons que la disponibilité et la flexibilité du SES, commanditaire de ces évaluations a également contribué à l'aboutissement du processus qui reste encore imparfait et devra être poursuivi tant il soulève des débats de fond.

Une caractéristique importante est le caractère ex post de ces évaluations qui portent sur des projets terminés depuis plus de quatre ans. Ceci permet un certain recul quant aux effets des projets sur les bénéficiaires et l'atteinte de l'impact. Cependant, le peu d'attention donnée par le dispositif de suivi-évaluation des projets aux *outcomes* et à l'impact, ajouté à la difficulté de remettre la main sur certaines données ou d'interroger certaines personnes des années après l'intervention, ont posé des obstacles à l'équipe empêchant ainsi l'application stricte de méthodes quantitatives.

Malgré ces difficultés, les évaluateurs ont réussi à utiliser certains outils quantitatifs. Une conception quasi expérimentale du processus d'évaluation a pu être mise en œuvre dans deux des quatre projets. Il a ainsi été possible de sélectionner a posteriori un contrefactuel crédible et de reconstituer des données « *baseline* » auprès d'un échantillon de bénéficiaires et d'un contrefactuel via des données de seconde main et/ou des données faisant appel à la mémoire des répondants. Cependant, étant donné les limites des données collectées et la crédibilité parfois remise en cause du contrefactuel, les outils économétriques n'ont pu être utilisés que pour démontrer l'attribution au projet d'une minorité des « *outcomes* ». Par contre, l'ensemble des projets évalués font l'objet d'un jugement argumenté sur l'atteinte des *outcomes* non traitables quantitativement, ainsi que sur l'atteinte de l'impact.

Remarquons que le contexte a changé depuis la mise en œuvre des quatre projets. Actuellement, des efforts sont fournis (notamment par la CTB) afin de mieux se centrer sur le suivi et l'évaluation des *outcomes* et de l'impact. Cependant, ces efforts doivent encore être mis en pratique et se traduire par des mesures concrètes et des

¹ Les méthodes quantitatives de type expérimentale préconisent d'assigner les personnes de manière aléatoire dans un groupe de bénéficiaires (groupe traité) et dans un groupe de non bénéficiaires (groupe de contrôle) par Essais contrôlés randomisés (ECR), plus connus sous sa terminologie anglaise : *Randomized Control Trial* (RCT).

changements sur le terrain. Il n'est pas garanti qu'il soit aujourd'hui plus aisé d'appliquer les méthodes quantitatives sur base des données collectées via le dispositif de suivi-évaluation.

Les évaluations de trois des quatre projets font l'objet de rapports séparés disponibles en annexe. Le rapport concernant l'évaluation du projet de laboratoire en santé publique au Rwanda n'est pas disponible car cette évaluation n'a pas pu être menée à son terme. L'équipe n'a pas reçu toutes les autorisations requises par la Ministre de la santé pour collecter et analyser les données. Le présent rapport synthétise donc les principales conclusions en termes d'atteinte des *outcomes* et de l'impact des trois autres projets. Les enseignements concernant la faisabilité et la valeur ajoutée d'une évaluation d'*outcomes* et d'impact s'inspirent par contre de l'ensemble des processus d'évaluation mis en œuvre dans les quatre pays.

Enseignements sur l'atteinte des *outcomes* et de l'impact des 4 projets évalués

Le projet d'appui à l'enseignement technique et professionnel en République démocratique du Congo (AETP2).

La méthodologie retenue pour cette évaluation est une étude de cas avec contrefactuel dans la ville de Kisangani. Cette étude se base donc sur une analyse qualitative rigoureuse des observations du terrain et des informations collectées par le biais d'entretiens individuels et de groupes menés dans des écoles bénéficiaires et des écoles témoins, ainsi qu'auprès de personnes clefs liées au secteur et/ou au projet en Belgique et en RDC. Des entretiens approfondis ont été menés auprès des directions d'école, des professeurs, des élèves et des anciens élèves diplômés. Etant donné l'absence de données quantitatives exploitables pour ce projet et vu que la mise en œuvre d'un dispositif de collecte de données de première main s'est révélée peu pertinente suite aux constats de la mission exploratoire, l'application de méthodes quantitatives a dû être abandonnée.

L'exercice réalisé dans les écoles de Kisangani démontre que **le projet n'a eu aucun effet significatif sur la qualité de l'enseignement, sur l'accroissement des effectifs, sur la capacité d'autofinancement des établissements scolaires, ni sur l'employabilité des jeunes diplômés.** Cet échec s'explique en partie par l'utilisation très partielle des « outputs », mais surtout par l'environnement très défavorable du secteur de l'enseignement technique et professionnel dans le pays. Depuis de nombreuses décennies les infrastructures se trouvent dans un état lamentable aggravé par les situations de conflits violents qui ont ravagé le pays. De plus, l'Etat n'a pas mis en place une politique adéquate soutenue par des moyens budgétaires appropriés (entre autres, l'école est majoritairement financée par les parents et leurs enfants). L'absence de systèmes incitatifs adéquats au sein des institutions et même l'existence d'incitants pervers nuisent à l'efficacité de toute action visant à améliorer la qualité de l'enseignement en RDC.

En réalité, dans les conditions mises en lumière qui étaient connues lors de la conception du projet et qui sont pour la plupart toujours d'actualité, ce projet n'aurait pas dû avoir lieu ou aurait dû être conçu différemment.

Projet d'amélioration et de renforcement des points d'eau dans le bassin arachidier au Sénégal (PARPEBA).

En l'absence de données « baseline » exploitables, un design quasi-expérimental a pu être mis en œuvre pour collecter des données de première main auprès de bénéficiaires ayant accès à l'eau du forage et auprès de ménages relativement similaires aux bénéficiaires mais n'ayant pas accès à l'eau du forage au sein de leur village. Il a donc été possible de comparer certains indicateurs d'*outcomes* entre deux groupes (groupe cible et groupe témoin) et entre deux périodes (pré- et post-projet) lorsque la reconstitution de la situation de référence était crédible via les données collectées par le biais d'enquêtes de ménages.

L'évaluation conclut que **l'approvisionnement en eau du forage via des robinets a permis un gain de temps et d'énergie significatif, spécialement pour les bénéficiaires ayant accès à un robinet privé**. Le temps et l'énergie gagnés n'ont apparemment pas eu d'effet sur la quantité d'eau consommée par jour et par personne. Par contre, ces gains semblent avoir encouragé le démarrage de petites activités commerciales, spécialement en saison sèche, ainsi que la participation dans des groupements et associations locales. De plus, il est observé que les usagers ont une forte disposition à payer pour l'eau du forage.

Un autre effet des nouveaux forages est **l'amélioration du bien-être et le renforcement de la cohésion sociale** (diminution des conflits entre les femmes, principales responsables de la collecte de l'eau). L'effet sur l'éducation mérite d'être approfondi car énormément de facteurs exogènes au projet peuvent expliquer l'évolution observée.

Néanmoins, **la qualité de l'eau tant au niveau bactériologique que chimique reste un réel problème**. Plus des deux tiers des sites construits ou réhabilités par le PARPEBA présentent des teneurs en fluor et en sel nocives pour l'organisme. De plus, les sources de contamination de l'eau entre le point d'approvisionnement et de consommation sont nombreuses. La population n'est pas sensibilisée à ces problématiques de qualité de l'eau. **Un tel constat appelle des mesures urgentes de correction au risque de causer des problèmes graves de santé aux populations rurales dans un avenir proche**.

Projet de petites et moyennes hydrauliques dans la région de Tiznit au Maroc (PMH).

Une conception quasi-expérimentale a également permis de comparer un groupe cible (agriculteurs bénéficiant des réhabilitations du système d'irrigation traditionnel) avec un groupe témoin (agriculteurs n'ayant pas bénéficié de ce type de réhabilitation) à deux moments dans le temps (pré- et post-projet) pour certains indicateurs d'« outcomes », grâce à la collecte de données de première main par le biais d'enquêtes de ménages.

L'évaluation montre que grâce aux opérations de réhabilitation du système d'irrigation de crues, **le projet a eu un effet positif sur les superficies cultivées en période de stress hydrique, alors qu'aucune amélioration n'est observée en période de crue**. Comparés aux agriculteurs non-bénéficiaires, les bénéficiaires du projet tirent une plus grande partie de leurs revenus de l'activité agricole non seulement en période de sécheresse mais aussi en période de crue. Cette dernière conclusion est à prendre avec précaution car elle se base sur une appréciation du revenu par les répondants eux-mêmes. En outre, le projet a encouragé l'extension spontanée des réseaux d'irrigation surtout dans les périmètres en amont. De plus, il a potentiellement permis de recharger la nappe phréatique.

Cependant, **le projet ne semble avoir encouragé ni la diversification des cultures ni des changements de pratiques culturales. Il n'a donc pas eu d'effet sur la diminution de l'exode rurale comme escompté**. Etant donné le caractère aléatoire des crues, ce projet mérite d'être suivi et évalué au cours de plusieurs années successives avec plus de précision concernant les rendements agricoles d'un groupe de bénéficiaires et d'un contrefactuel encore plus crédible. En effet, en dépit de ses faiblesses, il présente un grand intérêt pour la région.

Enseignements sur l'approche méthodologique

Cette étude confirme la littérature récente sur l'évaluation d'impact : **la combinaison de méthodes qualitatives et quantitatives au sein d'un même processus évaluatif est la solution adéquate pour réaliser des évaluations d'impact de qualité**. Dans le cas spécifique de cette évaluation, le processus évaluatif s'est basé sur la logique d'intervention du projet. Il valide donc la *theory based evaluation*. Cette approche utilise les méthodes quantitatives pour tenter, dans les limites de leur application, de mesurer les *outcomes* et d'en démontrer leur attribution au projet ; et les

méthodes qualitatives pour porter un jugement argumenté sur l'atteinte de l'impact, mais également sur l'atteinte des *outcomes* pour lesquels l'application des méthodes quantitatives n'est techniquement pas possible. Une analyse qualitative rigoureuse du contexte au sens large permet également de comprendre les mécanismes et facteurs ayant favorisé ou inhibé les effets d'un projet. **Un processus évaluatif combinant des outils quantitatifs et qualitatifs permet donc à la fois d'identifier 'ce qui a fonctionné' et 'ce qui n'a pas fonctionné', et de comprendre pourquoi et comment les effets ont été atteints ou n'ont pas été atteints.**

Bien qu'il soit pertinent de combiner les méthodes, les conditions pour utiliser de façon optimale les méthodes quantitatives n'étaient pas réunies pour l'ensemble des quatre projets évalués. **Les évaluateurs ont donc dû s'adapter aux limites imposées** par l'absence de « baseline », de données de suivi exploitables et de contrefactuel défini. Néanmoins, même si l'application des méthodes quantitatives a été fortement limitée, l'exercice a pu démontrer que ces méthodes s'enrichissent mutuellement.

Tout d'abord, nonobstant les contraintes, des outils quantitatifs d'analyse ont pu être utilisés pour quantifier certains *outcomes* et démontrer leur attribution. Ensuite, le dispositif de collecte de données mis en place pour l'application des méthodes quantitatives a généré un nombre plus important de données qualitatives grâce aux informations collectées lors des missions exploratoires, à la réalisation de collecte de données de première main via des enquêtes de ménages et aussi grâce à une présence relativement longue sur le terrain. Ces données sont exploitables pour argumenter l'atteinte des effets des projets (*outcomes* et impact) rendant l'analyse qualitative plus rigoureuse. Par ailleurs, une bonne analyse qualitative s'avère nécessaire à la mise en œuvre rigoureuse des méthodes quantitatives. Elle sert à argumenter des choix méthodologiques et à interpréter les résultats trouvés par les outils statistiques et économétriques. De plus, elle se prononce sur l'atteinte de l'impact alors que l'analyse quantitative se limite à l'atteinte des « outcomes ».

L'utilisation de méthodes mixtes permet également d'impliquer l'ensemble des acteurs liés à un projet : le dispositif quantitatif approfondit les entretiens avec les bénéficiaires, mais aussi avec des non-bénéficiaires ; le qualitatif permet quant à lui d'apprécier avec plus de précision le rôle de l'ensemble des acteurs gravitant autour d'une intervention.

Il existe cependant des conditions minimales requises pour que l'application de méthodes quantitatives soit pertinente. Dans une situation idéale, le projet aura via son dispositif de suivi-évaluation collecté des données fiables en quantité suffisante auprès du groupe bénéficiaire et d'un contrefactuel, avant et après l'intervention. Dans tout autre cas de figure, la faisabilité d'une analyse quantitative dépendra de l'identification d'un contrefactuel crédible, des possibilités de reconstruire des données pré-projet fiables, de l'aval du partenaire et des moyens disponibles. En effet, il ne faut pas négliger l'aspect politique d'une démonstration potentiellement négative des effets d'un projet. La rigueur des analyses quantitatives, peut induire les autorités locales à se montrer réticentes à soutenir l'évaluation. L'aval du partenaire est donc nécessaire pour mettre en œuvre un dispositif quantitatif d'évaluation.

Finalement, la question de la « value for money » de la mise en œuvre d'un tel processus de collecte et d'analyses quantitatives de données doit être abordée. Si une analyse qualitative rigoureuse démontre la faible probabilité d'observer un effet du projet, alors la mise en œuvre d'un dispositif quantitatif peut être remise en question, car une analyse quantitative n'apportera vraisemblablement rien de plus. Par contre, en cas de probabilité élevée de trouver un effet, la mise en œuvre d'un tel dispositif permettra de quantifier les effets sur les bénéficiaires et d'en démontrer les liens de causes à effets avec l'intervention, ce qu'une analyse qualitative ne sera pas en mesure de faire. **Plus fondamentalement, l'analyse coûts/bénéfices d'une évaluation d'impact utilisant des outils quantitatifs ne sera positive que lorsque les autorités belges en charge de la coopération au développement reconnaîtront la valeur ajoutée d'une analyse quantitative des « outcomes ».**

Par ailleurs, afin de limiter les coûts d'une analyse quantitative, plusieurs options sont possibles selon les objectifs poursuivis par l'évaluation : se limiter à étudier un aspect du projet, se concentrer sur une zone géographique, et/ou faire usage de données de seconde main fiables. Cependant, ne pas réaliser d'évaluation quantitative d' *outcomes* sous prétexte que les projets mis en œuvre sont calqués sur des projets dont leur succès a déjà été évalué et démontré dans d'autres études est très risqué. En effet, la validité externe des évaluations est souvent remise en question. De plus l'influence du contexte et du mécanisme d'intervention constituent deux éléments clés du succès d'un projet.

Enseignements en termes de conception/formulation, de mise en œuvre et de suivi-évaluation des projets

Les évaluations ont mis en évidence **des manquements au niveau de la conception, de la mise en œuvre et du suivi-évaluation des projets**. Les projets évalués présentent une logique d'action faible. Les objectifs sont déconnectés des activités et des moyens mis en œuvre. Ils sont peu précis et mal définis. L'objectif global est souvent irréaliste compte tenu du contexte local de l'intervention. Les faiblesses constatées sont préoccupantes dans la mesure où elles indiquent que l'atteinte des *outcomes* et de l'impact n'est pas une priorité dans les projets évalués. De tels manquements dans la conception montrent également que le cadre logique est réduit à une simple formalité. En effet, une analyse approfondie de la logique d'action aurait sans doute mené à une révision des quatre projets évalués. Ce pose ainsi la question de la pertinence dans le choix de certains projets.

Par ailleurs, le principe de l'alignement sur les politiques nationales pose également question lorsque celles-ci présentent des caractéristiques défavorables à l'atteinte des *outcomes* et de l'impact attendus. Plus fondamentalement, c'est la capacité de négociation avec le partenaire qui est mise en cause, ainsi que la capacité de la DGD et de la CTB à remettre en cause un projet peu réaliste.

La mise en œuvre privilégie le décaissement et les réalisations physiques. La gestion axée sur les « outputs » se fait au détriment d'une gestion axée sur les *outcomes* et l'impact. Ce mode de gestion est sans doute le reflet d'un manque d'incitants à faire autrement au sein du système de l'aide bilatérale belge.

Le système de suivi-évaluation des quatre projets est insuffisant et inadapté. Même s'ils sont conçus de façon pertinente, en pratique, les dispositifs de suivi-évaluation des quatre projets évalués n'ont permis ni le suivi ni l'évaluation des *outcomes* et de l'impact. En effet, bien que des indicateurs d'*outcomes* aient été définis ou cités dans la formulation du dispositif du suivi-évaluation, bon nombre n'a jamais été collecté et lorsque des données existent, elles ne sont pas exploitables quantitativement.

Recommandations

Il est recommandé de **promouvoir les évaluations d' *outcomes* et d'impact** car elles peuvent répondre adéquatement aux objectifs de reddition des comptes, d'aide à la décision et de capitalisation d'expériences (partage des leçons apprises). De plus, les constats d'une évaluation rigoureuse peuvent servir à différents acteurs : aux bailleurs, au partenaire, aux exécutants des projets et finalement aux bénéficiaires. La décision de mener des évaluations d'impact doit s'inscrire dans le contexte plus large du cycle de l'aide tenant compte des enjeux futurs de la coopération belge. Une stratégie évaluative doit également tenir compte du rapport coût/bénéfice qui restera faible tant que l'évaluation n'occupera pas une place plus importante au sein de la coopération belge.

La promotion de l'évaluation des *outcomes* et de l'impact, ainsi que son intégration dans une stratégie globale d'évaluation doit se faire en tenant compte de plusieurs éléments. Tout d'abord, il est nécessaire de clarifier les définitions d'*outcomes* et d'impact au moins au niveau des différents acteurs de la coopération belge. Ensuite, il faut encore faire connaître ce type d'évaluation auprès des partenaires, des personnes en charge de la conception/formulation et des exécutants des projets, afin de l'intégrer dans la gestion des projets. Il convient de prévoir un programme annuel d'évaluations des *outcomes* et de l'impact des interventions de la coopération gouvernementale belge en s'assurant de

leur valeur ajoutée ainsi que de leur pertinence, et en réalisant une étude de faisabilité au cas par cas.

Il est recommandé de réaliser des évaluations **combinant des méthodes quantitatives et qualitatives au sein d'un seul processus évaluatif**. Pour ce faire, il faut du temps et des moyens suffisants pour recruter une équipe maîtrisant à la fois les outils qualitatifs et quantitatifs d'analyse afin de garantir la qualité des évaluations. Il faut donc se donner les moyens de les réaliser avec la rigueur nécessaire et l'adaptation requise aux réalités du terrain.

Il est recommandé de **concevoir le dispositif d'évaluation des *outcomes* et de l'impact lors de la conception des projets**. Cela implique de revoir la conception, la mise en œuvre et le suivi-évaluation des projets afin que la poursuite des *outcomes* et de l'impact devienne une priorité au même titre que l'atteinte des « outputs ».

Concevoir l'évaluation dès les prémices d'un projet permet **d'utiliser le cadre logique comme véritable outil de formulation, de gestion axée sur les résultats et d'évaluation**. En effet, cela permet une prise en compte plus réelle des liens de causalité entre les activités, les moyens mis en œuvre et les objectifs fixés ainsi qu'une analyse de risques plus réaliste. Cela devrait, entre autres, avoir pour conséquence de ne plus valider les projets irréalistes. Concevoir le dispositif d'évaluation en détail avant une intervention nécessite de définir des indicateurs pertinents et d'en préciser les modalités de collecte avant, pendant et après l'intervention, ainsi que d'exposer les types d'analyses prévues. Par conséquent, si ce travail est bien fait au sein du projet, cela représente, au moment de l'évaluation des *outcomes* et de l'impact, un gain de temps et d'argent important, et permet également un saut significatif dans la qualité des évaluations.

Finalement pour être en mesure d'évaluer les *outcomes* et l'impact d'un projet par une méthodologie mixte, il convient **d'harmoniser la mise en œuvre du projet avec le dispositif d'évaluation**. La mise en œuvre séquentielle (« pipeline process ») est donc à considérer sérieusement afin de rendre possible à moindre frais et de façon plus éthique, l'utilisation d'un contrefactuel et donc l'application de méthodes quantitatives.

Toutes ces recommandations ne sont réalisables qu'en fournissant les **bons incitants** à chaque acteur impliqué dans la réalisation d'un projet de développement, tant du côté du pays donateur, que du pays bénéficiaire.

1. Introduction

L'objectif de ce rapport est de synthétiser les leçons apprises de l'évaluation d'impact ex post de quatre projets de la coopération gouvernementale commanditée par le Service de l'Évaluation spéciale (SES) du Service public fédéral belge des Affaires étrangères, Commerce extérieur et Coopération au Développement.

Cette évaluation a été réalisée dans quatre pays et secteurs différents par des experts du bureau d'étude ADE spécialisés en évaluation et des chercheurs de l'université de Namur (Centre de Recherche en Economie du Développement -CRED-). La collaboration entre un bureau de consultance et un centre de recherche académique a permis la mise en œuvre d'une approche méthodologique originale basée sur la logique d'intervention et utilisant tant les méthodes qualitatives que quantitatives. **Les enseignements portent sur l'atteinte des *outcomes* et de l'impact des projets étudiés, sur les méthodes utilisées, ainsi que sur la faisabilité et la valeur ajoutée de l'évaluation d'*outcomes* et de l'impact.** Le rapport amorce également une réflexion sur la place de l'évaluation d'impact dans le secteur de l'aide bilatérale belge et propose des pistes méthodologiques pour en réaliser.

Ces enseignements sont illustrés par des exemples tirés de l'évaluation de quatre interventions de la coopération gouvernementale belge: un projet d'accès à l'eau potable au Sénégal (PARPEBA, 2003-2008), un projet en petite et moyenne hydraulique au Maroc (PMH, 2004-2007), un projet de soutien à l'enseignement technique et professionnel en RDC (AETP2, 2005-2008) et finalement un projet d'appui à la mise en place d'un laboratoire de référence en santé publique au Rwanda (LRSP, 2003-2008).

Notons que l'évaluation du projet au Rwanda n'a pas été menée à son terme, faute d'accord pour la collecte de données indispensables pour l'analyse. En effet, les autorisations requises par la Ministre de la Santé en vue de la collecte et utilisation des données, ainsi que les réponses aux demandes d'autorisation n'ont pas été obtenues. En conséquence, le SES, en accord avec les consultants, a décidé de ne pas publier le rapport.

Il s'agit des premières évaluations d'impact (« impact evaluation ») stricto sensu menées par le SES. De plus, il s'agit d'**évaluations d'impact ex post**. Ces évaluations ont pour but de vérifier si ces projets terminés depuis plusieurs années ont atteint leurs objectifs (en termes d'*outcomes* et d'impact), et de comprendre les mécanismes menant aux constats de succès ou d'échec. Nonobstant les inévitables changements intervenus dans le fonctionnement de l'aide bilatérale belge depuis la conception et la mise en œuvre de ces projets -finalisés depuis 2007 ou 2008-, les conclusions restent pour la plupart pertinentes. Elles seront néanmoins nuancées lorsque cela s'avère nécessaire.

Ce rapport est structuré de la façon suivante. La section 2 soulève le débat sur la définition du terme 'impact' et présente les méthodes qualitatives et quantitatives dans le contexte actuel de l'évaluation d'impact. Ensuite, les objectifs et défis des méthodes quantitatives sont exposés. Pour finir, cette section résume l'approche méthodologique générale appliquée dans le cadre des quatre évaluations. La section 3 résume les enseignements sur l'atteinte des *outcomes* et de l'impact de trois des quatre projets évalués (comme déjà mentionné, l'évaluation au Rwanda n'a pas pu être menée à son terme) ; les rapports d'évaluation sont disponibles en annexes séparées. La section 4 tire les enseignements du processus d'évaluation combinant des méthodes qualitatives et quantitatives mises

en œuvre pour l'appréciation des *outcomes* et de l'impact des projets. Ce processus permet également de tirer des enseignements relatifs à la conception, la mise en œuvre et le suivi des projets. La section 5 présente les recommandations.

2. Méthodologie

2.1 L'impact, une définition à clarifier

L'absence d'une définition univoque de l'impact est une source d'incompréhension et de confusion qui nuit à une discussion sereine sur l'évaluation de l'impact et sa mise en œuvre.

Dans le contexte particulier du développement et de l'évaluation d'impact, coexistent principalement trois définitions :

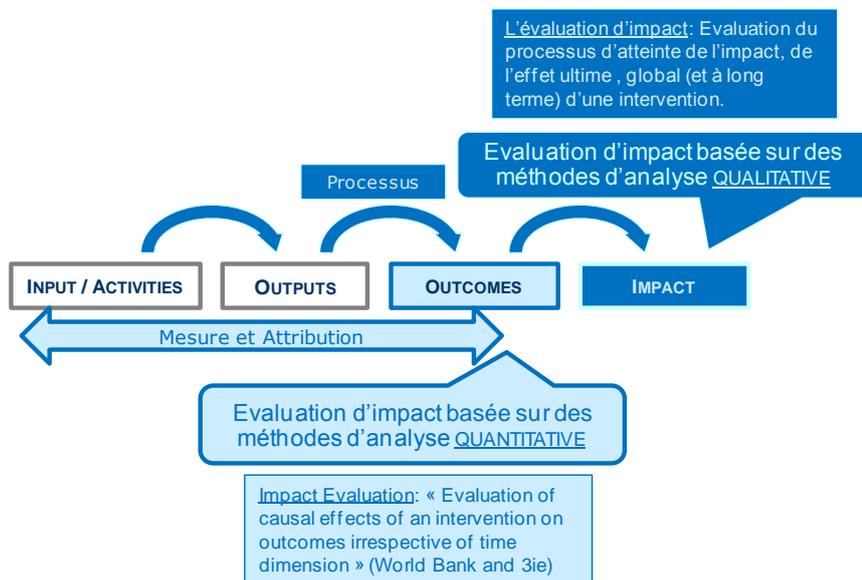
- (1) La définition de l'impact selon l'OCDE fait référence à la temporalité des effets d'une intervention. Dans le glossaire du **CAD**, l'impact est défini comme des **« effets à long terme positifs ou négatifs, primaires et secondaires, induits par une action de développement, directement ou non, intentionnellement ou non »**². La notion de temporalité amène une confusion quant aux effets à évaluer car le temps nécessaire à l'apparition des effets peut varier de façon significative en fonction du type d'intervention.
- (2) La deuxième définition est issue de la logique d'intervention telle que résumée dans le cadre logique où une action de développement est structurée en « input » (intrants/activités), « outputs » (résultats intermédiaires), objectifs spécifiques et objectif global. **L'impact selon l'approche de la logique d'intervention équivaut à l'objectif global, c'est-à-dire la contribution de l'intervention aux changements à un niveau global.** L'évaluation d'impact constitue donc à apprécier l'atteinte de l'objectif global. Notons que selon cette approche, les *outcomes* correspondent aux objectifs spécifiques.
- (3) Les **praticiens des méthodes quantitatives utilisent le terme impact pour désigner les « outcomes », qu'ils définissent comme les effets attribuables au projet sur les bénéficiaires, sans préoccupation de la temporalité de ceux-ci.** Ils définissent donc l'évaluation d'impact comme « *the evaluation of causal effects of an intervention on outcomes irrespective of time dimension* » (World Bank). La notion d'attribution fait référence à l'utilisation d'un contrefactuel (cf. infra), ce qui amène le groupe d'évaluation indépendante de la Banque Mondiale (IEG) à proposer une autre définition de l'évaluation d'impact : « *the impact evaluation is the counterfactual analysis of the impact of an intervention on final welfare outcomes* ». Le terme impact est donc utilisé alors qu'il fait clairement référence à l'évaluation des *outcomes* d'une intervention. Finalement, selon H. White « impact evaluation » est synonyme de « *attribution analysis* » ou encore « *with versus without analysis* » (*"what happened with the program -factual record- compared to what would have happened in the absence of the program -which requires a counterfactual-*", 3ie, 2013).

Comme illustré dans la [Figure 1](#), les méthodes qualitatives et quantitatives ne poursuivent pas les mêmes objectifs d'évaluation. Les premières s'intéressent à l'ensemble du processus menant à l'impact défini comme l'effet ultime et global d'une intervention. Les secondes mesurent les *outcomes* d'un projet, définis comme les effets sur les bénéficiaires directs, et en démontrent l'attribution.

² "Glossary of key terms in evaluation and results based management", OCDE – DAC (2010).

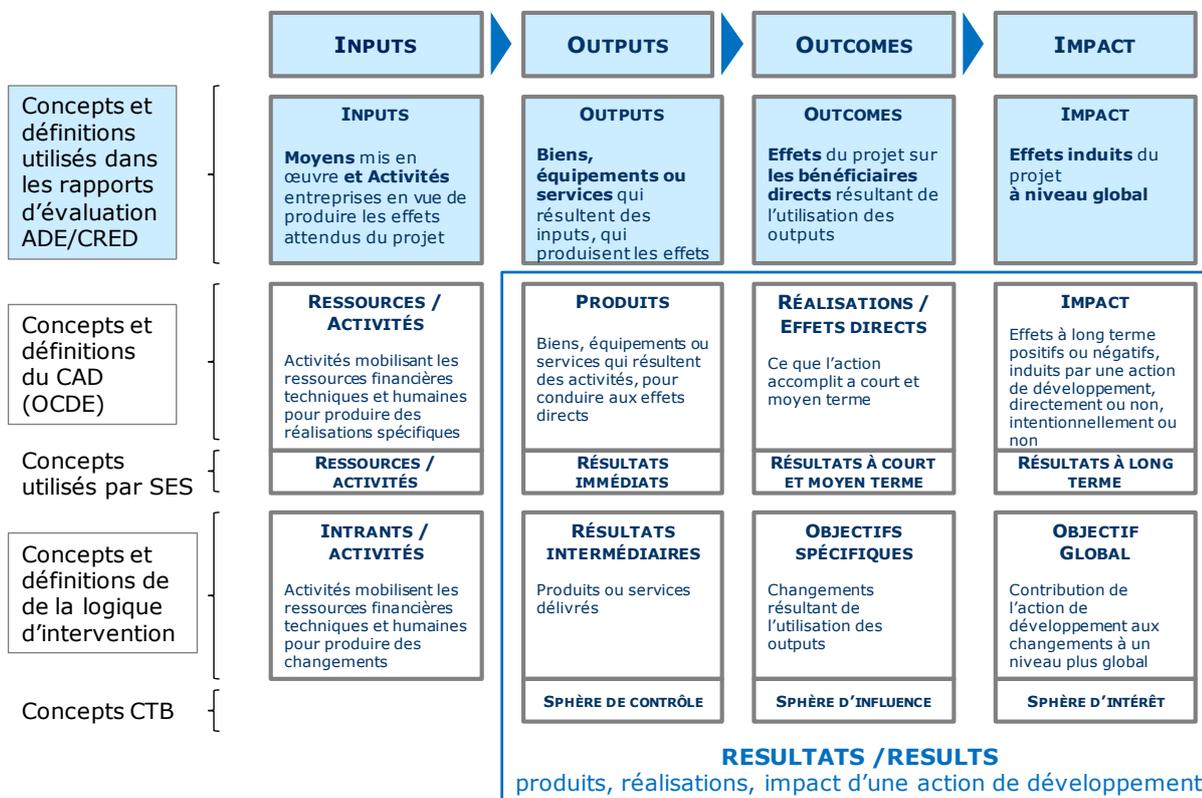
Les méthodes quantitatives n'ont pas pour objectif de se prononcer sur l'effet à un niveau global (impact) d'une intervention.

Figure 1 : Deux conceptions de l'évaluation d'impact



La Figure 2 présente les concepts utilisés dans le cadre logique avec les différentes définitions existantes. La première ligne de définitions correspond aux définitions utilisées dans le cadre des présentes évaluations. Les deux autres lignes exposent respectivement les définitions du CAD et de la logique d'intervention.

Figure 2 : Différentes définitions des concepts du cadre logique



Dans le cadre des quatre évaluations d'impact réalisées, l'approche de la logique d'intervention a été retenue. Cela signifie (i) que les définitions d'*outcomes* et d'impact ne font pas référence à la temporalité des effets ; et (ii) que l'évaluation d'impact des projets s'articule autour du cadre logique en combinant deux méthodes : une quantitative pour apprécier et démontrer l'attribution des effets sur les bénéficiaires directs³ (« *outcomes* ») lorsque cela est techniquement possible, et qualitative pour apprécier l'effet à un niveau global (impact) et apprécier les *outcomes* lorsque l'application des méthodes quantitatives posent problème. Cela équivaut à dire que l'on se réfère à la « *theory based evaluation* » qui utilise à la fois des méthodes quantitatives et qualitatives.⁴

Outre les différences au niveau de la définition du terme 'impact', la figure 2 indique qu'il existe également différentes traductions françaises d'un même concept du cadre logique. C'est pour cette raison que le présent rapport, ainsi que les 3 rapports d'évaluation disponibles en annexe, utilisent les termes en anglais afin d'éviter toute confusion supplémentaire.

Par ailleurs, la Figure 2 indique que le terme « résultats » désigne tout ce qui résulte des inputs, c'est-à-dire les « *ouputs* », les *outcomes* et l'impact. Cette observation a toute son importance lorsqu'on parle de « *results based management* » ou de « gestion axée sur les résultats ». En effet, le dispositif de suivi-évaluation doit être en mesure d'apprécier à la fois les « *ouputs* », les *outcomes* et l'impact d'un projet.

2.2 Un processus évaluatif, deux méthodes

Comme évoqué au point précédent, **la méthodologie retenue pour évaluer les *outcomes* et l'impact est une combinaison de méthodes qualitatives et de méthodes quantitatives articulées autour de la logique d'intervention.** Ces méthodes sont brièvement présentées ci-dessous.

Les méthodes qualitatives émettent un jugement argumenté sur l'atteinte de l'impact d'une intervention, entendu comme l'effet au niveau global d'un projet (cf. Figure 1 ci-dessus). Elles analysent si les conditions pour atteindre l'impact souhaité sont remplies aux différentes étapes de l'intervention, en référence au cadre logique: conception/formulation, input, mise en œuvre, réalisation et utilisation des « *ouputs* », atteinte des *outcomes* et atteinte de l'impact.

Les analyses qualitatives se basent sur des informations collectées via une analyse documentaire, des entretiens individuels ou de groupe, et par des observations directes sur le terrain. Une bonne analyse qualitative apporte une compréhension approfondie du contexte au moment de l'intervention. En plus de situer le projet dans une perspective historique (projets antérieurs, prévus, etc.), cette démarche consiste à comprendre et identifier tous les aspects socio-économiques, politiques et culturels, ainsi que les enjeux de pouvoir à différents niveaux qui peuvent avoir eu une influence sur le projet et/ou sur le bon déroulement de l'évaluation.

Les méthodes quantitatives mesurent et démontrent l'attribution des *outcomes* d'une intervention, définis comme les effets sur les bénéficiaires directs qui sont attribuables au projet. L'objectif poursuivi est d'établir un lien de causalité en démontrant empiriquement dans quelle mesure un projet *-et uniquement ce projet-* a

³ Une analyse quantitative ou qualitative peut également apprécier les effets d'une intervention sur des bénéficiaires indirects lorsque cela s'avère pertinent.

⁴ La « *theory based evaluation* » est souvent assimilée aux méthodes qualitatives, car elle se base sur l'analyse du cadre logique, peu utilisé par les praticiens des méthodes quantitatives (ils évaluent souvent un aspect d'un projet dans une perspective de recherche et pas l'ensemble du projet). La « *theory based evaluation* » n'est pas contradictoire avec l'utilisation de méthodes quantitatives, mais elle n'en fait pas toujours systématiquement l'usage.

2. Méthodologie

contribué aux effets observés sur les bénéficiaires (« outcomes »). Cela consiste à utiliser des techniques statistiques et économétriques sur des données fiables en quantité suffisante collectées auprès d'un groupe de bénéficiaires et d'un contrefactuel, avant et après l'intervention (cf. infra). Ces données peuvent être collectées directement sur le terrain via des enquêtes de ménages et/ou être des données de seconde main, c'est-à-dire des données existantes provenant de sources diverses (disponibles via le dispositif de suivi-évaluation des projets, via les statistiques nationales, etc.).

Notons que l'utilisation des méthodes quantitatives dans l'évaluation de projet s'inscrit généralement dans une perspective de recherche où la conception du projet est concomitante à la conception du dispositif d'évaluation, permettant ainsi de réaliser l'évaluation dans les meilleures conditions techniques.⁵ En effet, la faisabilité d'appliquer des méthodes quantitatives dépend fortement de la quantité et de la qualité des données disponibles avant et après l'intervention, ainsi que de la possibilité d'identifier un contrefactuel crédible. Les défis et limites de l'application de ces méthodes sont nombreux, particulièrement dans un contexte d'évaluation ex post (cf. infra).

Bien que l'analyse qualitative soit en mesure de porter un jugement argumenté sur l'atteinte des *outcomes* d'un projet, l'application de méthodes quantitatives sera privilégiée sur l'analyse qualitative lorsque cela s'avère techniquement possible. En effet, dans les conditions idéales, l'application stricte des méthodes quantitatives permet de quantifier les *outcomes* et d'en démontrer scientifiquement la causalité avec le projet (son attribution), avec un certain degré de précision. Cependant, si les conditions nécessaires pour la bonne application des outils économétriques ne sont pas toutes remplies, l'analyse quantitative ne sera pas en mesure de démontrer l'attribution, mais seulement un niveau de corrélation entre les effets observés et le projet. Les liens de causalité devront alors être argumentés par l'analyse qualitative. Par ailleurs, si l'application des méthodes quantitatives se révèle impossible par manque de données ou encore à cause de biais trop importants, une analyse qualitative rigoureuse sur l'atteinte des *outcomes* est alors souhaitable.

Concernant l'appréciation de l'effet à un niveau global (impact) d'une intervention, les méthodes quantitatives sont peu efficaces, là où les méthodes d'analyse qualitative sont performantes. Par contre, les constats sur les *outcomes* faits par l'analyse quantitative peuvent être utilisés comme arguments forts dans l'appréciation de l'impact et de son ampleur. Parfois ils peuvent être généralisés ou extrapolés, s'ils sont accompagnés d'une explication convaincante.

Ces méthodes bien que très différentes ont besoin l'une de l'autre pour fournir des évaluations d'*outcomes* et d'impact de qualité. Une parfaite maîtrise des techniques quantitatives ne pourra jamais se substituer à une analyse rigoureuse de la chaîne de causalité d'une intervention, ni à une argumentation pertinente réalisée par des experts connaissant le terrain (Deaton, 2012). Inversement, une histoire convaincante pour expliquer les effets d'une intervention ne remplacera pas la mesure et la preuve scientifique de l'attribution des effets d'un projet grâce à l'utilisation de méthodes quantitatives.

Finalement, une analyse d'*outcomes* et d'impact de qualité ne se limite pas à identifier ce qui a fonctionné ou ce qui n'a pas fonctionné mais doit expliquer comment et pourquoi ces résultats ont été obtenus ou n'ont pas été obtenus. Une analyse qualitative

⁵ Ces méthodes sont nées et ont été diffusées plus largement par J-Pal (Abdoul Latif Jamel Poverty Action lab) ; elles ont été depuis utilisées par de nombreux chercheurs universitaires pour réaliser des évaluations d'impact de projets financés par divers bailleurs (Banque Mondiale, AFD, Coopération Hollandaise, coopération Norvégienne, etc.), mais à chaque fois, la formulation et la mise en œuvre des projets à évaluer ont été réalisées en cohérence avec un dispositif d'évaluation assurant les meilleures conditions d'applicabilité des techniques quantitatives. Le présent exercice est différent car ce sont des évaluations ex post d'impact de projet pour lesquels aucun dispositif d'évaluation d'impact n'existait, l'application des méthodes quantitatives restent donc limitée.

rigoureuse du contexte et des enjeux à différents niveaux, ainsi qu'une bonne connaissance des réalités du terrain sont donc nécessaires pour expliquer les résultats obtenus via l'application de techniques statistiques et/ou économétriques.

En résumé, la combinaison de méthodes qualitatives et quantitatives permet une évaluation précise des *outcomes* au moyen d'un contrefactuel crédible ainsi qu'une évaluation de l'impact par une analyse factuelle rigoureuse. En d'autres termes, une méthodologie mixte permet une mesure des effets sur les bénéficiaires et une démonstration de leur attribution ou de leur corrélation avec le projet (selon la quantité de données fiables disponibles et l'utilisation possible d'un contrefactuel pour la bonne application des méthodes quantitatives). Elle permet aussi un jugement argumenté sur l'atteinte de l'impact d'un projet, basé sur certains constats prouvés scientifiquement grâce à l'analyse quantitative des « *outcomes* ». Finalement, elle permet d'apporter une explication du processus qui a généré ou inhibé les effets d'un projet.

2.3 Méthodes quantitatives : objectifs et défis

Les méthodes quantitatives ont pour objectifs (1) d'identifier et de mesurer les changements observés dans la population cible depuis la mise en œuvre du projet ; et (2) d'établir, avec un certain degré de précision, un lien de causalité entre les changements observés (attendus ou imprévus) et le projet.

1) Le premier objectif des méthodes quantitatives est de mesurer les effets d'un projet sur les bénéficiaires (« *outcomes* »).

Une fois que les *outcomes* de l'intervention ont été identifiés dans l'analyse du cadre logique, il convient de se poser une série de questions :

- a. Quels sont les *outcomes* mesurables ?
- b. Comment peut-on les mesurer ?
- c. A-t-on les données disponibles pour les mesurer ?
- d. Comment collecter des données lorsqu'elles ne sont pas disponibles ?

Soulignons que si les données n'existent pas et qu'il n'y a pas moyen de les collecter ou si un « *outcome* » n'est pas mesurable, il n'est alors pas possible d'utiliser des méthodes quantitatives. Par exemple au **Maroc**, il s'est avéré que mesurer la productivité des terres était un indicateur pertinent mais trop complexe à collecter avec les moyens et le temps à disposition. Cet indicateur a donc été abandonné. Dans les trois évaluations, les indicateurs d'*outcomes* ont été construits sur base des informations collectées sur le terrain mais tous n'ont pas pu faire l'objet d'analyses quantitatives, soit parce qu'ils se sont avérés peu fiables ou parce qu'il n'y avait pas suffisamment d'observations. La qualité et la quantité des données sont deux éléments clés pour l'application des méthodes quantitatives (cf. infra). Notons qu'une variable qualitative peut être traitée avec des méthodes quantitatives, il suffit qu'elle soit collectée de façon systématique avec un nombre suffisamment large d'observations pour être en mesure de la traiter statistiquement ou de l'utiliser dans une régression économétrique⁶.

⁶ Exemples de questions qualitatives qui peuvent être traitées statistiquement : « *depuis l'arrivée de l'eau du forage dans votre village estimez-vous que vos conditions de vie se sont améliorées ?* » OUI/NON ; « *depuis l'arrivée de l'eau du forage dans votre village, situez sur une échelle de 1 à 5 le gain de bien-être pour vous et votre famille ? (1 : pas de gain ; 5 : gain maximum)* » 1-2-3-4-5.

2) Le second objectif des méthodes quantitatives est de démontrer l'attribution des effets à l'intervention.

Afin de conclure à une relation de causalité, il faut démontrer que les changements observés sur les bénéficiaires résultent de l'intervention.⁷ En d'autres mots, il faut pouvoir estimer qu'elle aurait été la situation de la population soumise au projet en l'absence de celui-ci. L'idéal serait donc de comparer les mêmes individus avec ou sans mise en œuvre du projet, ce qui est par définition impossible. Une option est donc de comparer la population soumise au projet (groupe traité) à une population qui lui est similaire selon un nombre important de caractéristiques préalables à l'intervention. Ce groupe s'appelle groupe de contrôle, groupe témoin ou encore contrefactuel.

Pour atteindre ces deux objectifs, **les méthodes quantitatives font face à plusieurs défis.**

1. Le choix d'un bon contrefactuel

Le groupe de traitement, encore appelé groupe cible, correspond à la population bénéficiaire sélectionnée pour recevoir l'intervention de développement. Le contrefactuel est une population non-bénéficiaire non significativement différente de la population traitée, avant l'intervention. En d'autres termes, les deux groupes sont similaires sur un nombre important de caractéristiques, idéalement, la seule différence serait que le premier soit soumis au traitement et l'autre pas.

Deux types de conception (« design ») d'évaluation existent pour déterminer le contrefactuel : une conception expérimentale et une conception quasi-expérimentale.

- Considérée comme la plus solide des méthodes d'évaluation, « **the experimental design** », encore appelé « **the randomized control trial** » (**RCT**) est la première façon de sélectionner un contrefactuel. Le fondement de cette méthode est d'affecter de manière aléatoire le projet parmi des bénéficiaires éligibles d'une même population, de façon à générer un groupe de traitement et un contrefactuel équivalents statistiquement avant l'intervention. Cette sélection se fait souvent à un certain niveau selon le type d'intervention, comme par exemple au niveau des villages ou des écoles. Lorsque la sélection se fait au niveau des individus, cela peut poser certaines questions éthiques.⁸ Avec un nombre suffisamment important d'observations, le processus d'assignement aléatoire permet de constituer deux groupes dont les caractéristiques moyennes sont statistiquement équivalentes avant que le projet ne débute. L'utilisation de l'assignement aléatoire garantit donc, en théorie, que ces groupes sont équivalents. La collecte de données de base permet de vérifier empiriquement cette hypothèse. Si ces deux groupes sont effectivement statistiquement équivalents avant la mise en œuvre du projet, alors une simple différence ex post entre le résultat moyen constaté pour le groupe de traitement et le résultat moyen observé pour le groupe de contrôle suffira.

⁷ Parfois, les spécifications économétriques possibles avec les données disponibles présentent des biais d'endogénéité rendant l'analyse de la causalité complexe ou impossible. Dans ces cas-là, les résultats obtenus dans les régressions sont des corrélations entre l'« outcome » analysé et le projet. Rien techniquement ne permet d'affirmer que ces effets sont dus à l'intervention, ils pourraient être expliqués au moins en partie par des facteurs exogènes non observables.

⁸ Voir White (2013) « An introduction to the use of randomized control trials to evaluate development interventions », pour plus de détails sur l'application des RCT pour l'évaluation d'impact de projets. Voir aussi Getler et al. (2011) « L'évaluation d'impact en pratique » pour une explication simple de l'assignement aléatoire des groupes.

Les avantages de cette conception sont : (i) le caractère aléatoire du choix des sujets permet d'éviter le problème du biais de sélection (cf. infra); (ii) la simplicité dans l'interprétation de la mesure des effets attribués au projet (par la méthode de la double différence -cf. infra pour une explication de l'utilisation de cette méthode-).

Les principaux inconvénients sont les suivants : (i) il peut être éthiquement ou politiquement difficilement acceptable de faire bénéficier un groupe et pas un autre d'une intervention ; (ii) la portée d'un projet peut ne pas permettre la constitution de deux groupes (par exemple, un projet à portée nationale) ; (iii) le caractère aléatoire peut être difficile à respecter ou les individus des deux groupes peuvent se déplacer au cours de l'intervention (« *spillover effect* » cf. infra) ; et (iv) c'est une méthode coûteuse qui demande beaucoup de temps tant pour la collecte que pour le traitement d'un nombre très important de données.

- Une **conception quasi-expérimentale** signifie que le groupe de bénéficiaires et le contrefactuel ne sont pas déterminés de façon aléatoire. Il s'agit de déterminer un contrefactuel présentant des caractéristiques observables statistiquement équivalentes au groupe bénéficiaire de l'intervention. Cela peut se faire avant ou après la mise en œuvre du projet.

Les avantages d'un tel « *design* » sont (i) qu'il est souvent moins coûteux et plus rapide qu'un « *design* » expérimental, et (ii) qu'il peut être mis en œuvre pour une évaluation ex post. L'inconvénient majeur est que les méthodes statistiques et économétriques à utiliser pour démontrer l'attribution des effets au projet sont souvent plus complexes, car il faut s'assurer qu'il n'y ait pas de biais de sélection (cf. infra). La fiabilité des résultats est donc parfois remise en question. Des sous-estimations ou surestimations des effets du projet sont possibles.

Deux problèmes fréquents sont à écarter pour démontrer que l'on a trouvé un bon contrefactuel : le problème de contagion, encore appelé de problème de contamination ou « *spill over effect* », et le problème du biais de sélection des bénéficiaires.

- **L'effet de contagion** se décline de deux façons exposées ci-après. Il est possible de surmonter ce problème avec une connaissance approfondie du contexte tant au niveau des dynamiques locales que de la présence d'autres bailleurs sur le terrain.
 - (a) Un effet indirect de l'intervention sur le contrefactuel. Pour assurer une certaine similitude entre les deux groupes, on aurait tendance à prendre des groupes relativement proches d'un point de vue géographique. Or, un groupe de contrôle voisin du groupe de traitement a de fortes chances d'être indirectement touché par le projet. Ainsi, même s'ils sont similaires avant l'intervention, les données post-intervention collectées auprès du groupe témoin sont influencées par le projet. Ceci fausse l'interprétation des analyses. Par ailleurs, un effet de contagion est un effet indirect du projet. Dans l'évaluation d'impact de l'accès à l'eau d'un forage au **Sénégal**, un effet de ce type a été observé : les ménages des villages voisins non-bénéficiaires s'approvisionnent aujourd'hui au robinet des villages bénéficiaires du projet.
 - (b) Effets d'autres projets sur le contrefactuel. Il est possible que le même genre d'intervention soit mené par d'autres bailleurs auprès du groupe de contrôle. Il est également possible que des interventions différentes soient conduites mais qu'elles affectent les mêmes « *outcomes* ». Dans ces deux cas de figure, une comparaison en termes d'*outcomes* entre le groupe cible et le groupe contrôle sera biaisée.
- **Le biais de sélection** apparaît lorsque les raisons pour lesquels une personne participe à un projet sont corrélées aux résultats. Ce concept fait donc référence

2. Méthodologie

à la façon dont les individus sont sélectionnés pour pouvoir bénéficier d'un projet. Si la participation au projet est déterminée sur base de caractéristiques observables des bénéficiaires (sexe, âge, analphabétisme, etc.), des techniques économétriques relativement simples existent pour surmonter le problème de biais de sélection. Cependant, il est souvent argumenté que des caractéristiques non-observables (ou difficilement observables), influencent la participation des bénéficiaires à un projet (« *self-selection problem* »). Si ces caractéristiques sont corrélées aux *outcomes* du projet, alors les analyses sont biaisées (biais d'endogénéité). En effet, des différences entre groupe bénéficiaire et contrefactuel pourraient être expliquées par les caractéristiques non observables et non pas uniquement par le fait de bénéficier du projet.⁹

Une façon de surmonter ce biais est l'assignation aléatoire du projet au sein de la population (« *randomized control trial* », cf. supra). Une autre façon de surmonter ce problème est d'utiliser la technique du « **pipeline** ». Cela consiste à mettre en œuvre un projet de façon séquentielle et d'utiliser comme contrefactuel le groupe de bénéficiaires faisant partie d'une phase d'exécution ultérieure. Cependant, l'ordre d'exécution du projet dans les diverses zones d'intervention ne peut pas être décidé sur base de caractéristiques influencées par le projet ou influençant les *outcomes* du projet ; il est donc préférable de choisir un ordre aléatoire d'exécution dans la mesure du possible.

Pour démontrer la qualité du contrefactuel choisi, il est important d'apporter des justifications convaincantes, limitant toute suspicion de biais de sélection. Il convient de démontrer que le groupe cible et le contrefactuel ne sont pas statistiquement différents avant la mise en œuvre de l'intervention, d'où l'importance de disposer de données « baseline » au sein des deux groupes. Si cela n'est techniquement pas possible (manque de données), il convient néanmoins d'argumenter sur base d'une bonne connaissance du terrain, qu'il n'y a pas de biais de sélection ou au minimum identifier la direction du biais : une surestimation ou une sous-estimation des effets.

Les évaluations ex post des projets au Maroc et au Sénégal sont des conceptions quasi expérimentales dans la mesure où aucun contrefactuel n'a été défini préalablement, d'une part, et d'autre part, aucune donnée « baseline » n'est disponible. Les contrefactuels identifiés ne sont pas tout à fait satisfaisants mais l'analyse le reconnaît et en tient compte. Au **Maroc**, les villages témoins présentent des caractéristiques pré-projet différentes des villages cibles. Ces caractéristiques étaient vraisemblablement observables (entre autres le type d'agriculture) mais n'ont malheureusement pas été collectées faute de connaissance du contexte. L'interprétation des résultats est donc prudente et tient tant que possible compte de ce biais. Les résultats non interprétables ont été abandonnés. Au **Sénégal**, les villages traités et témoins présentent également des différences pré-intervention. Grâce à l'existence de diagnostics participatifs qui présentent un état de la situation dans les villages avant l'intervention, certaines différences observables ont pu être identifiées et intégrées dans l'analyse pour identifier la direction du biais (exemples de caractéristiques observables au niveau du village : densité de population, nombre d'utilisateurs par points d'eau, enclavement...). Par ailleurs, un problème d'auto-sélection existe concernant les utilisateurs de borne privée. L'accès à un robinet dans la cour est un effet indirect du projet, et la sélection des bénéficiaires ayant accès à un robinet privé reste en partie méconnue par manque de variables observables

⁹ L'exemple suivant illustre ce propos. Dans un projet de microcrédits, on serait tenté de comparer en termes de revenus générés (« outcome ») des individus avec et sans microcrédit. Cependant, il est fort probable que les individus les plus dynamiques fassent la demande d'avoir accès à un microcrédit. Cette caractéristique difficilement observable détermine donc le fait d'avoir accès au projet et la façon de faire fructifier son argent. Une comparaison d'individus avec et sans microcrédit donnera vraisemblablement une surestimation de l'effet d'avoir accès à un microcrédit.

collectées, mais aussi à cause de la grande influence possible de variables non-observables.

2. L'utilisation de données de qualité en quantité suffisante

Les méthodes statistiques et économétriques utilisées pour estimer les *outcomes* d'une intervention nécessitent des données fiables en quantité suffisante. Avoir des données en quantité suffisante signifie disposer d'un nombre d'observations suffisamment large pour appliquer des techniques statistiques et économétriques (plus le nombre d'observations est élevé, plus la précision des tests est grande); mais également pour être en mesure de détecter les effets, le cas échéant. Avoir des données de qualité signifie que ces données soient pertinentes, qu'elles aient été définies de façon claire et précise, et qu'elles aient été collectées avec soin. On n'insistera jamais assez sur l'importance de la définition d'un indicateur, et sur les nombreux biais de mesure et de rapportage qui peuvent exister au moment de la collecte des données. Satisfaire ces deux conditions, la qualité et la quantité, reste un défi dans la plupart des évaluations, spécialement si l'on veut appliquer des méthodes quantitatives.

Compromis (« Trade-off ») quantité/qualité des données. Plus le nombre d'observations, tirées de façon aléatoire, est élevé, meilleure sera l'analyse statistique. En théorie, il existe une formule pour déterminer la taille optimale d'un échantillon permettant d'affirmer avec un certain degré de précision (« *power calculation* ») qu'il existe ou non un impact¹⁰. En pratique¹¹ cette formule reste en toile de fond, la taille de l'échantillon étant principalement calculée en fonction de la méthode statistique-économétrique envisagée et en fonction des contraintes temporelles et budgétaires. Par contre, il est inutile d'avoir une quantité importante d'observations si la fiabilité des informations collectées n'est pas bonne ou si les informations collectées sont inutiles. Il est primordial de ne pas négliger la qualité des données.

Les données peuvent être de deux types : des données de seconde main ou de première main.

Diverses sources de données de seconde main existent :

- des données du projet (via le dispositif de suivi-évaluation),
- des données provenant des statistiques nationales,
- des données collectées au sein d'autres projets,
- des données disponibles dans certains organismes publics ou privés (registres de santé, registres de présences scolaires, données au sein d'entreprises, etc.).

Trouver des données de seconde main de qualité est un réel défi dans de nombreux pays. Au **Sénégal**, il a été envisagé de collecter des données de seconde main auprès des écoles et des centres de santé (pour estimer l'accroissement de la présence à l'école et la diminution des diarrhées), mais l'idée fut vite abandonnée, étant donné la qualité très hétérogène des données existantes. Les registres de santé ou scolaires étant remplis au bon vouloir des personnes en charge ne constituent pas une source de données fiables. En **RDC**, les taux de réussite aux examens d'Etat ont été collectés via les directeurs d'école, ainsi que le nombre d'inscriptions dans les différentes filières. Au **Rwanda**, le cas est un peu particulier : des données de seconde main de qualité satisfaisante étaient disponibles pour étudier un « *outcome* » en particulier (l'amélioration du diagnostic du test de la goutte épaisse) mais elles n'ont pas pu être utilisées faute d'autorisation du Ministère de la santé rwandais.

Lorsqu'aucune base de données de seconde main n'est exploitable ou que celles-ci sont incomplètes ou peu fiables, il convient de collecter des données de première main.

¹⁰ Cette formule dépend de la taille de l'échantillon (plus large, plus le degré de précision des analyses sera élevé), de la corrélation intra cluster, de l'effet minimum (plus l'effet attendu est faible, plus large devra être l'échantillon), l'intervalle de confiance visé (95% ou 90% ou 99%) et des coûts d'enquête. Lire White (2013), « An introduction to the use of randomized control trials to evaluate development interventions ».

¹¹ C'est-à-dire lorsque la perspective de recherche académique n'est pas prioritaire.

Les données de première main sont des données directement collectées par l'équipe d'évaluation via des questionnaires détaillés adressés à des répondants pertinents pour le type d'étude envisagé. Par exemples : des ménages, un individu du ménage, des chefs de villages, des maires, des chefs d'entreprise, des directeurs d'hôpitaux ou d'écoles, etc. Ces enquêtes peuvent se faire à différents niveaux pour une même étude. Par exemple, au **Sénégal**, une enquête au niveau des ménages a été réalisée (217 ménages ont été interrogés), ainsi qu'une enquête au niveau des villages et des groupes d'usagers des forages. Bien que ces deux derniers types d'enquête soient plutôt des entretiens de groupe (« focus group », cf. infra), les données collectées ont été introduites dans la base de données ménages, afin de capter des différences entre villages ou entre groupes de gestion de forage. Au **Maroc**, une enquête de ménage a également été réalisée. En **RDC**, des données de première main ont été collectées à différents niveaux via des entretiens de groupes (« focus group ») : directeurs, professeurs, élèves et anciens élèves diplômés. Étant donné le petit nombre d'écoles sélectionnées (deux écoles bénéficiaires et deux écoles témoins), aucun traitement statistique approfondi n'a été réalisé.

Une collecte de données de première main nécessite une connaissance approfondie du contexte et des enjeux afin de créer de bons outils de collecte. Il convient de poser des questions pertinentes pour les thématiques étudiées dans le contexte local, mais il convient également de bien formuler les questions en utilisant un vocabulaire adapté aux types d'interlocuteurs recherchés. Un dispositif de collecte de données de première main fait appel à une mobilisation importante de ressources sur le terrain, et demande une implication des évaluateurs à chaque étape pour assurer la fiabilité des données. Cette implication va de la conception des questionnaires, à la supervision quotidienne de la collecte de données sur le terrain.

- **Les enquêtes type « focus group »** sont des entretiens structurés/semi-structurés avec un groupe plus ou moins important d'individus d'un même type (par exemple, les autorités villageoises, les gestionnaires d'un forage, un groupe de professeurs ou encore un groupe d'élèves). La mise en œuvre d'un tel outil demande une grande attention afin de détecter des incohérences ou des rétentions d'informations par l'un ou l'autre des participants. Bien que **cette technique relève des méthodes qualitatives**, elle est citée ici car les informations collectées par ce biais de façon systématique auprès de différents groupes peuvent être encodées dans une base de données et donc être traitées à l'aide d'outils statistiques (comme ce fut le cas pour le projet au **Sénégal**, les données collectées via les « focus group » avec les autorités villageoises ont été encodées).

Soulignons que pour une plus grande pertinence des données collectées, il est important que ces groupes soient sélectionnés aléatoirement. Cela s'applique également lorsqu'il faut sélectionner certains individus dans un groupe. En d'autres termes, pour garder son indépendance d'évaluateur, il est primordial « d'aller où on veut aller et pas où l'on nous suggère d'aller ». De plus, il faut veiller à ce qu'il n'y ait pas d'observateur pouvant inhiber certains participants. Par exemple, en **RDC**, les préfets n'étaient pas présents lors des entretiens avec les élèves ou les professeurs ; au **Sénégal**, aucun gestionnaire des réseaux de distribution de l'eau n'était présent lors des enquêtes avec les ménages bénéficiaires ou non-bénéficiaires.

- **Les enquêtes de ménage (« Household Survey »)** sont des entretiens basés sur un questionnaire détaillé administré par un enquêteur à un ou plusieurs individus d'un ménage, d'une entreprise, d'une école, etc. Dans le cas d'une évaluation, ces individus sont sélectionnés aléatoirement au sein d'une population de bénéficiaires et au sein d'une population de non bénéficiaires (contrefactuel). L'avantage de ce type d'outil est le fait que les répondants se sentent plus à l'aise pour parler que dans un entretien en groupe, les données collectées portent sur des informations quantitatives (par exemple, combien de personnes résident dans le ménages, quel est l'âge du chef de famille) mais également qualitatives (par exemple, quel est votre niveau de satisfaction de l'utilisation des robinets publics,

quelle est l'utilité d'avoir un diplôme secondaire pour avoir un travail en RDC...). Ces questionnaires posent également des questions, qu'il convient ensuite de codifier pour que les réponses soient traitables statistiquement.

La mise en œuvre d'une enquête de ménages nécessite une organisation logistique importante. Il faut constituer une équipe d'enquêteurs honnêtes et travailleurs. Il convient de tester le questionnaire pour s'assurer de la pertinence des questions et de leur compréhension, tant pour les répondants que pour les enquêteurs. Ensuite, durant la période d'enquête, un suivi rapproché des enquêteurs est nécessaire pour s'assurer de la qualité des données collectées mais également pour s'assurer que les données ont bien été collectées auprès des ménages présélectionnés (de façon aléatoire). La saisie de ce type de données demande également une supervision importante afin de s'assurer une fois de plus de la fiabilité des informations encodées qui seront ensuite analysées.¹² Finalement, il faut 'nettoyer' les bases de données. Il s'agit de procéder à divers tests afin de s'assurer que les données aient été encodées correctement et que les réponses soient cohérentes. Cela signifie qu'il n'y ait pas d'incohérence entre les réponses à diverses questions au sein d'un questionnaire et que l'ensemble des réponses soit cohérent avec les observations réalisées sur le terrain. Par exemple, pour la base de données du **Sénégal**, une erreur dans le canevas informatique n'a pas été détectée par le superviseur : cela a décalé toutes les réponses d'une ou deux cases dans une section. Il a fallu tout ré-encoder. Les consultants s'en sont rendu compte car certaines réponses n'étaient pas cohérentes avec ce qui avait été observé sur le terrain ; les données encodées ont donc été vérifiées avec celles des questionnaires originaux pour identifier la source de l'erreur. Sans cette prise de conscience, les réponses d'une question auraient été attribuées à une autre question et ainsi de suite.

Cette logistique représente un travail conséquent. Collaborer avec un partenaire local maîtrisant les techniques d'évaluation quantitatives, et qui a un réseau et de l'expérience dans ce type d'entreprise facilite énormément le travail sur le terrain.

La mise en œuvre de la collecte de données de première main engendre une série de coûts fixes (ne dépendant pas du nombre d'enquêtes ou d'entretiens à réaliser). Par exemple, quel que soit le nombre d'enquêtes à réaliser, il faut rédiger un questionnaire, former des enquêteurs et/ou un traducteur, tester les outils de collecte sur le terrain, faire des corrections éventuelles, concevoir un masque de saisie informatique pour encoder les données, etc. Réaliser des économies au cours de ces étapes cruciales du processus d'évaluation engendre inévitablement une perte de la qualité des données à analyser. Cela peut alors remettre en question la pertinence de la collecte de données de première main (cf. supra : compromis quantité/qualité).

Quel que soit le type de collecte de données, les acteurs observés peuvent avoir tendance à changer leurs comportements parce qu'ils se savent observés. Ceci s'appelle **« the hawthorne effect »**. C'est un problème typique de validité interne des évaluations d'impact. Les répondants obéissant à cet effet 'psychologique' ont une propension plus importante à déguiser la réalité (surestimer ou sous-estimer certains faits ou données). Il est donc important de rester vigilant face aux différents enjeux et stratégies mises en place par les répondants lors des évaluations, et tenter de capter les réponses qui seraient sous ou surestimées. La qualité du questionnaire, l'honnêteté des enquêteurs et le temps passé à gagner la confiance des autorités locales et de chaque répondant sont autant d'éléments qui peuvent fortement atténuer cet effet. Au **Maroc**, les autorités locales ont vraisemblablement influencé les répondants tantôt dans un

¹² Notons qu'il est de plus en plus fréquent d'observer des enquêtes réalisées sur des tablettes informatiques. La saisie des données est alors instantanée. Cependant l'utilisation d'un tel support comporte certains inconvénients : nécessité d'avoir accès à l'électricité, nécessité de trouver des enquêteurs familiers avec l'informatique, difficulté de revenir en arrière si l'on réalise une incohérence au cours de l'enquête, etc.

sens, tantôt dans l'autre. Heureusement, grâce à une longue présence sur le terrain qui a permis de gagner la confiance de la majorité des répondants, ce phénomène a pu être identifié au moins en partie.

Pour démontrer l'attribution, il est nécessaire **d'avoir ces données pour plusieurs périodes clefs de l'intervention collectées au sein du groupe de bénéficiaires et du contrefactuel**. L'idéal est de disposer de données pour la période pré-intervention (« **baseline** »), pendant l'intervention (« **monitoring** ») et pour la période post-intervention (« **ex post** »).

- **L'inexistence d'une enquête de base** peut entraîner un biais dans l'estimation des *outcomes*: (i) les différences observées via une enquête ex post entre un groupe cible et groupe témoin peuvent être préexistantes au projet ; et à l'inverse, (ii) une absence de différence entre ces deux groupes peut simplement refléter une différence préalable au projet, qui a disparu grâce à la présence de l'intervention dans le groupe cible. Une comparaison ex post entre participants et non participants à un projet soulève également le problème de biais de sélection (cf. supra : les raisons pour lesquelles une personne participe à un programme sont corrélées aux résultats).

Sans « baseline », il est possible d'utiliser des méthodes quantitatives sur des données ex post, pour comparer groupe cible et groupe témoin. La technique du « *Propensity score matching* » est la plus utilisée dans ce cas-là (cf. infra pour des explications). Cependant, sans « baseline », la crédibilité du contrefactuel doit être prouvée par une argumentation qualitative rigoureuse et les risques de biais dans les estimations sont plus grands (risque de comparer des groupes pas comparables, risque de ne pas contrôler pour des variables non observables influençant le résultat et le fait d'appartenir à un groupe).¹³ «Une fois le programme mis en œuvre, s'il n'est pas possible d'identifier avec précision la manière dont il a été attribué aux bénéficiaires, et qu'aucune donnée de base n'a été collectée, il reste peu, voire pas, de possibilité d'évaluation quantitative fiable » (Getler et al., 2011 : p115).

Lorsqu'il n'existe pas de « baseline », il convient donc, lorsque cela est possible, de reconstruire des données de base par le biais de données faisant appel à la mémoire des répondants (données « *recall* », cf. infra) ou via des données de seconde main fiables.

- **L'inexistence d'un contrefactuel**. Sans groupe de comparaison, les effets d'un projet sont déterminés en examinant l'évolution des résultats pour les participants au projet au fil du temps. Cela consiste à réaliser une comparaison avant-après des bénéficiaires sous hypothèse que sans le projet, les caractéristiques du groupe des bénéficiaires n'auraient pas évolué. Malheureusement dans la majorité des cas, cette hypothèse n'est pas valable. Sans groupe de contrôle valide, il est donc impossible de démontrer empiriquement que les effets observés au sein d'un groupe de bénéficiaires sont le résultat de l'intervention.
- Lorsque **les contraintes de temps et de budget sont serrées**, une option satisfaisante permettant l'application des méthodes quantitatives est de collecter des données de base auprès d'un échantillon de bénéficiaires, et de reconstruire ces données pour le contrefactuel via des données « *recall* » en ex post (cf. « *Conducting impact evaluations under budget, time and data constraints* », IEG, 2013).
- **Les données « recall »** sont des données collectées auprès des individus en faisant appel à leur mémoire afin qu'ils donnent les informations les plus précises possibles sur une situation antérieure à la période de l'enquête (par exemple : la

¹³ Voir Getler et al. « L'évaluation d'impact en pratique » p.114-115.

situation pré-intervention). Il est ici nécessaire d'avoir une référence dans le temps qui soit comprise par tous de la même manière. Or, la notion du temps est quelque chose de très culturel. Il est donc souvent utile d'utiliser des références historiques pour remettre les individus dans la situation voulue (par exemple : lors de l'élection de tel président, au moment de la sécheresse de telle année, avant la construction du robinet public, etc.). Certains types d'informations se prêtent mieux que d'autres à cette technique. En effet, des données exceptionnelles sont plus susceptibles d'être retenues avec précision par la mémoire que des informations concernant des faits quotidiens. Par exemple dans le cas du **Maroc**, il est plus facile de se souvenir des cultures réalisées en période de crue que de se souvenir des montants gagnés par l'activité agricole. En effet, la période de crue a lieu en moyenne tous les deux ans alors que les activités agricoles ont lieu tous les ans, toute l'année.

3. L'application rigoureuse des techniques statistiques et économétriques.

Pour démontrer empiriquement les liens de causalité entre les effets observés et le programme de développement, il convient d'utiliser des modèles économétriques plus ou moins sophistiqués selon les cas de figure. Les éléments suivants influencent le choix des techniques : conception ex ante ou ex post de l'évaluation, assignation aléatoire de l'intervention, existence de données de base, existence ou niveau de validité d'un groupe de contrôle, types de données, nombre de données et leur fiabilité. Cependant, il ne suffit pas de maîtriser la théorie, il faut maîtriser l'application de ces modèles à des bases de données concrètes. Il faut ensuite être capable d'interpréter les résultats en cohérence avec les hypothèses qui sous-tendent les modèles économétriques mais également en cohérence avec les réalités du terrain.

Dans les cas du **Maroc** et du **Sénégal**, aucune donnée pré-intervention n'était exploitable. L'équipe a tout même réussi à appliquer des méthodes quantitatives dans le but de quantifier et de démontrer l'attribution de certains « outcomes ». Deux techniques ont principalement été utilisées sur des données collectées ex post auprès d'un groupe de bénéficiaires et d'un contrefactuel : **(i) la méthode de la double différence** pour laquelle des données de base ont été reconstituées grâce à des données de « recall », et **(ii) la technique de l'appariement communément appelé en anglais « propensity score matching » (PSM)** pour laquelle l'appariement est principalement fait sur base de données ex post (ce qui n'est pas l'idéal, cf. supra). La validité du contrefactuel est démontrée empiriquement lorsque cela s'est avéré possible, sinon sa validité est principalement argumentée sur base d'observations.

Des analyses statistiques ont également été réalisées : des **tests de différences de moyennes ou d'indépendance de facteurs**. Ci-dessous, ces différentes techniques d'analyse sont brièvement expliquées (cf. « *Handbook on impact evaluation, quantitative methods and practices* », Khandker et al. (2010) et « L'évaluation d'impact en pratique », Getler et al. (2011), pour plus de détails sur ces méthodes et des applications concrètes).

❖ **Test de différence de moyennes**

Ce test permet de vérifier si deux groupes sont en moyenne statistiquement différents selon certaines caractéristiques en données continues (par exemple : âge du chef de famille, surface de terre possédée, nombre d'années d'éducation, etc.).

L'hypothèse nulle du test statistique est que les deux groupes sont en moyenne statistiquement similaires (H_0 : moyenne du groupe cible = moyenne du groupe témoin). Le test 'student' permet de conclure que l'on peut rejeter l'hypothèse nulle (c.-à-d. rejeter le fait qu'ils soient en moyenne statistiquement similaires), avec une certaine précision quant aux chances de se tromper : 10%* de chances de se tromper, 5%** ou 1%*** de chance de se tromper. Plus il y a d'étoiles, moins il y a de chances de se tromper en rejetant l'hypothèse nulle, c'est-à-dire plus on est en mesure d'affirmer que ces groupes sont en moyenne différents l'un de l'autre.

2. Méthodologie

Exemple :

Caractéristiques continues	Groupe cible	Groupe témoin	Significativité
Âge moyen du chef de famille	58.4 ans	57.6 ans	Pas significatif
Nombre moyen d'années d'éducation	10.8 années	9.5 années	**
Surface moyenne des terres possédées	9.13 ha	15.93 ha	***

Lecture des résultats de ce tableau :

- En moyenne, les chefs de famille des deux groupes ont des âges non significativement différents.
- En moyenne, les individus des ménages cibles ont plus d'une année supplémentaire d'éducation que les individus des ménages témoins, cette différence est significative à 5% (c.-à-d. qu'il existe 5% de chance que cette différence ne soit pas significative).
- En moyenne, les ménages cibles possèdent 5.8 ha de moins que les ménages témoins, cette différence est significative à 1% (c.-à-d. qu'il existe 1% de chance que cette différence ne soit pas significative).

❖ **Test d'indépendance de facteurs**

Ce test chi-carré teste l'hypothèse que deux variables discrètes (binaires ou pouvant prendre plusieurs valeurs) sont indépendantes l'une de l'autre. Une des variables est dichotomique indiquant si les observations appartiennent au groupe de contrôle ou au groupe témoin, l'autre représente une caractéristique discrète par exemple le fait d'avoir un scooter, une moto, un vélo ou une voiture ; le fait d'appartenir à une certaine ethnie (Wolof, Peulh, Serer).

L'hypothèse nulle indique que les variables sont indépendantes les unes des autres. Par exemple, l'hypothèse nulle stipule que le fait d'appartenir à une certaine ethnie n'est pas lié au fait d'être dans le groupe cible ou le groupe de contrôle. Le nombre d'étoiles indique que l'on peut rejeter, avec une certaine probabilité, le fait que ces variables soient indépendantes l'une de l'autre (cf. test de différence de moyenne, ci-dessus). Par exemple, plus il y a d'étoiles, plus on est confiant que le fait d'appartenir à une ethnie est lié au fait de faire partie d'un groupe (témoin ou cible).

Exemple :

Variable discrète : ethnie	Groupe cible	Groupe témoin	Significativité
Wolof	60%	40%	***
Peulh	12%	29%	***
Serer	16%	22%	Pas significatif

Lecture des résultats de ce tableau :

- La proportion de ménages wolof dans le groupe cible est significativement plus importante que dans le groupe témoin. Les deux groupes diffèrent donc de façon significative sur cette caractéristique. Rejet de l'indépendance des facteurs à 1% (c.-à-d. qu'il existe 1% de chance de se tromper en affirmant qu'il y a significativement plus de wolof dans le groupe cible)
- Le fait d'appartenir à l'ethnie Serer ne semble pas lié au fait d'appartenir à un certain groupe. Il y a donc non rejet de l'indépendance des facteurs.

❖ La double différence ou différence-en-différence

L'effet du traitement est déterminé en prenant la différence en termes d'indicateurs d'intérêt entre le groupe de traitement et le groupe de contrôle avant et après la mise en œuvre du projet (le traitement). Une variation significative à travers le temps de la différence en termes d'*outcomes* entre les deux groupes est alors attribuée au projet.

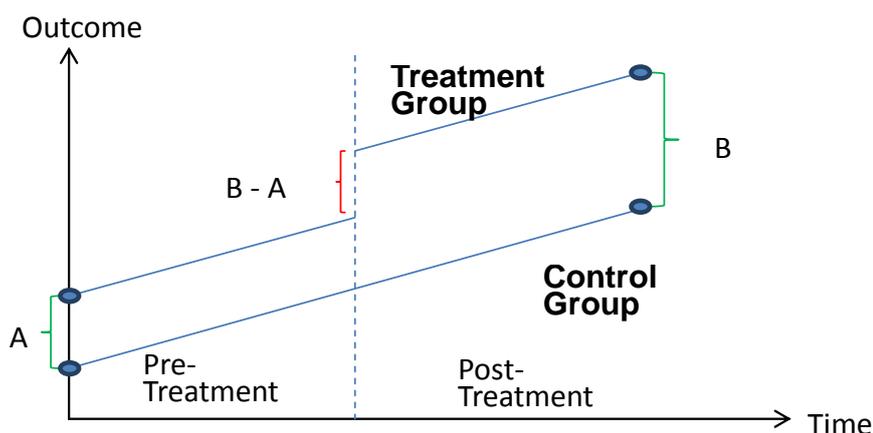
Etape 1: Collecte de données pour une année de référence avant l'implémentation du projet (ou données de base reconstruites via des données « *recall* » avec toute la prudence requise en termes de fiabilité des données) ; et observation de la différence entre les deux groupes (traitement et contrôle) avant le programme ;

Etape 2: Collecte de données pour une année de référence postérieure à l'implémentation du projet ; et observation de la différence entre les deux groupes (traitement et contrôle) après le projet ;

Etape 3: Observation de l'évolution de la différence entre les deux groupes (la méthode se focalise sur la différence entre les différences). La méthode de la différence-en-différence est basée sur les **hypothèses** suivantes:

- Les caractéristiques non-observables ne varient pas à travers le temps : le biais de sélection, s'il existe, est constant à travers le temps et les caractéristiques non-observables qui pourraient influencer l'« outcome » ne changent pas à travers le temps.
- Les deux groupes suivent la même tendance en l'absence d'un traitement.

Figure 3 : Technique de la double différence schématisée



Exemple :

Type de ménages	Ménages cibles	Ménages Témoins	Différence entre type de ménage à une période
Périodes			
Pré-projet (2006)	132 minutes	150 minutes	18 minutes
Post-projet (2012)	59 minutes	124 minutes	65 minutes ***
Différence pré/post projet au sein d'un groupe	73 minutes ***	26 minutes	Diff-en-diff : 47 minutes ***

Lecture du résultat du tableau : Grâce au traitement (l'arrivée d'un robinet public), les ménages cibles ont un gain de temps de 47 minutes par jour.

❖ **Technique d'appariement (« Propensity score matching »)**

La méthode vise à comparer une série de variables d'intérêt (qui mesurent l'effet potentiel du projet) entre individus qui ont une probabilité identique de participer au projet (cette probabilité peut être calculée de multiples façons).

Etape 1 : Création d'un contrefactuel valide en appariant les ménages sur base de caractéristiques observables (non influencées par le projet) - idéalement sur des données de base- ;

Etape 2 : Sur base des caractéristiques observables, la probabilité de participation au traitement est estimée pour tous les ménages ;

Etape 3 : Chaque ménage du groupe de traitement est comparé à des ménages du groupe cible selon une règle à définir. Par exemple à ces 5 'voisins' les plus proches en termes de probabilité de participation (méthode de calcul utilisée dans l'évaluation Maroc) ;

Etape 4 : Le gain dû au traitement est calculé par la différence entre l'« outcome » du ménage traité et l'« outcome » moyen de ses 5 plus proches 'voisins' (ou selon la règle choisie à l'étape 3) ;

Etape 5 : Le résultat obtenu par cette méthode est le gain moyen total du traitement. Il est donné par la moyenne des gains individuels calculé en étape 4.

La méthode du PSM est basée sur l'**hypothèse** qu'il n'existe pas de caractéristiques non-observables susceptibles de conjointement expliquer l'effet étudié du projet et le fait d'être dans le groupe de traitement (pas de biais de sélection). Le grand problème avec cette méthode c'est qu'il n'est pas possible de prouver empiriquement qu'il n'existe pas de caractéristiques non-observables susceptibles d'influencer la participation au projet et le résultat. De plus, cette technique nécessite de collecter de grandes bases de données sur des échantillons importants. En outre, il existe toujours un risque de manque de support commun entre les deux groupes (cible et témoin). Notons également qu'un appariement sur des données ex post n'est pas sans risque. On peut involontairement effectuer un appariement sur base de caractéristiques qui ont été affectées par le programme d'aide.

Exemple :

Variabes	Coefficient	Significativité	Nbr. d'obs.
Part moyenne des revenus agricoles dans les revenus totaux	14	*** (1%)	218

Lecture du résultat de ce tableau :

La méthode du PSM conclut qu'il existe un effet significatif (à 1%) du traitement sur la proportion du revenu total provenant de l'activité agricole. Cet effet est positif la proportion des revenus agricoles atteint près de 40% des revenus totaux pour le groupe de traitement contre 26% pour le groupe de contrôle (différence de 14 point de pourcentage est le coefficient donné par l'estimation).

4. **L'interprétation correcte et cohérente des résultats**

Même si la lecture des résultats est facile une fois que l'on maîtrise les techniques quantitatives, l'interprétation peut poser quelques difficultés. D'une part, l'application de deux modèles différents peut ne pas donner les mêmes résultats. Il convient donc soit de continuer les analyses pour trouver un résultat robuste¹⁴, soit de rester prudent quant à l'existence de ce résultat car il n'est pas robuste. D'autre part, les résultats obtenus peuvent être contraires aux observations réalisées sur le terrain. Il convient alors de chercher à comprendre ce résultat en s'assurant que les méthodes ont bien été appliquées, que les données sont correctes et que les observations du terrain sont convaincantes. S'il existe une divergence entre un résultat quantitatif et une analyse qualitative, il convient de le notifier et de rester prudent quant à l'interprétation de ce résultat.

De plus, il faut rester prudent lors de la conclusion sur l'existence ou non d'un « outcome » lié à une intervention de développement. En effet, plusieurs éléments peuvent biaiser les résultats obtenus. Tout d'abord, la petite taille des échantillons peut réduire la précision des estimations rendant discutable l'identification ou non d'un effet. Ensuite, il faut s'assurer de la bonne utilisation des méthodes économétriques. Pour cela, il faut faire des estimations qui ont du sens tant techniquement que contextuellement et prendre en compte les biais potentiels (biais de sélection, biais de contagion). Finalement, la fiabilité des données est essentielle, il faut un maximum se prémunir d'erreurs de mesures et des stratégies de manipulation des réponses par les répondants. Par ailleurs, il faut, en tant que lecteur, toujours rester critiques quant aux possibles manipulations des chiffres par les praticiens (« *How to lie with statistics* », Darrell, 1993). La restitution des résultats sur le terrain est souvent un bon moyen d'éviter ce genre de manipulation et donc une interprétation erronée des résultats empiriques.

L'interprétation des résultats nécessite donc une bonne connaissance technique, une honnêteté intellectuelle mais également une bonne compréhension de la thématique du projet et du contexte local. Il est dès lors plus facile pour un évaluateur d'interpréter les résultats d'une étude lorsqu'il a, à la fois géré le processus de collecte de données du début à la fin, et été en charge de l'analyse. C'est le cas pour les deux évaluations d'impact ex post réalisées au **Sénégal** et au **Maroc** : le chercheur du CRED a été présent tout au long du processus.

Dans le cas du **Maroc**, seule étude où les deux techniques ont été testées, l'équipe a été confrontée au fait que ces deux techniques ne donnaient pas toujours le même résultat. Certes, ceci signifie que ces résultats ne sont pas robustes (fiables), mais cela illustre aussi la difficulté inhérente à l'interprétation des résultats, la prise en compte de biais potentiels, ainsi que l'honnêteté et la prudence requise dans l'interprétation des résultats.

¹⁴ Un résultat robuste signifie un résultat qui se maintient à travers divers modèles et sous différentes spécifications. Concrètement, cela signifie qu'un résultat soit significatif, de relativement la même ampleur, et du même sens (positif ou négatif) lorsqu'on applique par exemple un modèle de différence-en-différence et un modèle de PSM ; et lorsque pour un même modèle, on change certaines variables explicatives.

2.4 L'organisation pratique de l'évaluation

Les évaluations d'impact publiées qui utilisent des méthodes mixtes ont souvent été exécutées par des équipes spécialisées dans une de ces méthodes. Elles ont été réalisées :

- Soit par des praticiens des méthodes qualitatives qui utilisent à certains moments des méthodes quantitatives. Cependant, vu la complexité de mise en œuvre des méthodes quantitatives, ces évaluations manquent souvent de rigueur à plusieurs niveaux.
- Soit par des praticiens des méthodes quantitatives (principalement des personnes issues du monde de la recherche académique) qui emploient des méthodes qualitatives pour justifier certains choix méthodologiques et interpréter les résultats. Néanmoins, les méthodes qualitatives sont souvent reconnues par ces derniers comme un « second best ».

La particularité des évaluations d'impact menées ici réside dans le fait que chaque évaluation ait été réalisée par deux équipes travaillant en parallèle, chacune spécialisée dans un type de méthode : ADE pour l'application des méthodes qualitatives et le CRED pour l'application des méthodes quantitatives.

Le cas de la **RDC** est un peu particulier car le processus d'évaluation utilise uniquement des méthodes qualitatives mais sur base d'un processus évaluatif quantitatif. En effet, c'est une étude de cas basée sur des entretiens individuels et de groupes. Cependant, le choix de la province, mais également le choix des répondants a été fait de la façon la plus aléatoire possible et une attention particulière a été donnée à interroger les bénéficiaires directs de l'intervention. De plus, une approche avec un contrefactuel a été retenue et le temps passé sur le terrain était plus long que pour une étude qualitative classique.

Bien que les processus d'évaluation aient été mis en œuvre de façon relativement indépendante par les deux équipes, un rapport d'évaluation commun par évaluation a été produit. L'idée était d'enrichir mutuellement les constats sur les *outcomes* et l'impact sans *a priori* défini dans le processus d'évaluation à suivre. La mise en commun des résultats pour aboutir à un seul rapport de qualité par projet, ainsi que les échanges avec le comité d'accompagnement furent des moments riches d'enseignements.

La Figure 4 ci-dessous reprend **les étapes clés suivies par chacune des équipes puis celles du travail de mise en commun.**

Figure 4 : Etapes clés du processus d'évaluation par des méthodes mixtes

	ADE « méthodes quali »	CRED « méthodes quanti »
Phase préparatoire	Reconstruction du cadre logique	Identification des <i>outcomes</i> réalistes et mesurables Identification de la stratégie méthodologique Mission exploratoire
Phase terrain	Confrontation des relations causales aux réalités du terrain	Collecte de données de 1 ^{ère} main Saisie des données
Phase d'analyses	Jugement argumenté sur l'atteinte de l' impact	Nettoyage et traitement des données Mesure et démonstration de l'attribution des outcomes

Mise en commun	
Etape 1	Confrontation des logiques d'intervention reconstituées
Etape 2	Confrontation des appréciations des <i>outcomes</i> et impacts; explications des mécanismes
Etape 3	Rédaction d'un rapport commun

Pour **l'équipe ADE**, spécialisée dans l'application des méthodes qualitatives, voici les étapes successivement mises en œuvre :

- **Phase préparatoire : La reconstruction de la logique d'intervention du projet** sur base d'une étude documentaire et d'une mission sur le terrain. Ce travail permet de mettre en évidence la chaîne de résultats escomptés qui, par liens de cause à effet, devrait produire l'impact attendu.
- **Phase terrain : La confrontation des relations causales aux réalités du terrain**, par le biais d'observations des réalisations et de leur utilisation, d'avis d'experts, d'entretiens semi-structurés avec diverses parties prenantes tant en Belgique que dans le pays d'intervention (dont certains bénéficiaires). Les experts se sont également appuyés sur une analyse avec contrefactuel même si celle-ci n'est pas statistiquement représentative (c'est le cas en **RDC**).
- **Phase d'analyse : L'appréciation de l'impact à un niveau global** basée sur un jugement argumenté de la contribution de l'intervention à l'impact (à un niveau global, cf. définition supra). Les méthodes d'analyse qualitative sont généralement reconnues comme étant peu efficaces pour quantifier l'impact avec précision et adresser son attribution au projet. Cependant, leur force réside dans leur capacité à éclairer la contribution potentielle du projet à l'impact dans des contextes complexes.

Pour **l'équipe du CRED**, spécialisée dans l'application de méthodes quantitatives, le processus d'évaluation est brièvement décrit ci-dessous.

- **Phase préparatoire :**
 - **Identification des *outcomes* réalistes et mesurables de l'intervention** sur base d'une lecture des documents du projet disponibles mais également sur base de la littérature académique existante sur les thématiques traitées dans le projet et d'autres évaluations d'impact réalisées dans les mêmes domaines/secteurs.
 - **Identification de la stratégie méthodologique.** Cela consiste à identifier des méthodes quantitatives pertinentes pour quantifier et attribuer les effets (« *outcomes* ») au projet ; et à identifier une stratégie de collecte de données après avoir dressé la liste des données déjà existantes qui seraient exploitables pour une analyse quantitative. Cette étape consiste à définir l'échantillon, le contrefactuel, les outils de collecte de donnée à utiliser, etc.
 - **Conception des outils de collecte de données de première main** (questionnaire de ménages, guide d'entretien des « focus-group ») grâce à une bonne connaissance du contexte, mais aussi du projet et des *outcomes* à mesurer.
 - **Mission exploratoire.** Cette mission poursuit plusieurs objectifs.
 - (i) Elle a pour but de mieux connaître le contexte, les enjeux et les réalités du terrain en rencontrant diverses personnes clefs du projet et du secteur d'intervention, ainsi qu'en réalisant une visite dans la zone d'intervention.
 - (ii) Elle sert à vérifier un certain nombre de points cruciaux pour l'analyse. Cela consiste à vérifier la pertinence du choix des *outcomes* à mesurer, par conséquent, il faut commencer par observer la réalisation des « *outputs* ». Il faut aussi vérifier la fiabilité des données de seconde main disponibles et s'assurer de la faisabilité de la stratégie de collecte de données de première main envisagée.
 - (iii) C'est l'occasion de rencontrer le partenaire local pour organiser la logistique de la phase terrain (constitution d'une équipe d'enquêteurs, arrangements financiers, déplacements sur le terrain...) et affiner la méthodologie sur base de ses connaissances du terrain.
 - (iv) Finalement, c'est l'occasion de tester les outils d'enquête et d'y apporter des corrections.

- **Phase terrain** :
 - **Collecte de données de première main.** Cette étape commence par la formation des enquêteurs et par un test du questionnaire auprès d'un échantillon de bénéficiaires et de non-bénéficiaires. Ensuite, les enquêtes se déroulent sur le terrain avec une supervision rapprochée de l'évaluateur pour garantir la fiabilité des données collectées.
 - **Saisie des données** sous un format informatique (par exemple sur le programme CPro), supervisée par le partenaire local (l'encodage des données est un processus lent et donc coûteux dans notre pays, il est préférable de le réaliser sur place).
- **Phase d'analyse** :
 - **Nettoyage des données** pour vérifier la qualité de l'encodage (saisie cohérente avec les données présentes dans les questionnaires papier).
 - **Traitement des données par l'application de méthodes quantitatives** (par exemple avec le logiciel STATA). Les hypothèses sous-jacentes aux effets du projet sont testées par des techniques statistiques et économétriques.
 - Sur base des résultats des analyses, **l'équipe se prononce sur l'atteinte des « outcomes », les quantifie et en démontre leur attribution**, le cas échéant.

Le travail de mise en commun des résultats des deux équipes s'est fait en fin de processus.

- **Etape 1 : Confrontation de la logique d'intervention** explicitement reconstituée par ADE avec la logique d'intervention implicitement reconstituée par le CRED lors de l'identification des *outcomes* à apprécier.¹⁵
- **Etape 2 : Confrontation des appréciations des *outcomes* et de l'impact** pour chaque projet. A ce stade, plusieurs éléments découverts par le biais de l'une ou l'autre méthode servent d'arguments supplémentaires aux constats réalisés par chaque équipe. Certains constats sont également remis en cause. C'est ici, que l'approche mixte prend tout son sens.
- **Etape 3 : Rédaction d'un rapport d'évaluation commun.**

¹⁵ Les chercheurs du CRED, comme beaucoup d'académiques n'utilisent pas la logique d'intervention. Cependant, pour identifier les « outcomes » mesurables, ils en ont reconstruite une sans le savoir. C'est pour cela que le terme « implicitement » est utilisé. Dans l'avenir, cette étape devrait se faire conjointement, afin de gagner du temps.

3. Principales conclusions sur l'atteinte des *outcomes* et de l'impact des 3 projets

Cette section présente dans un premier temps un résumé des conclusions relatives à l'atteinte des *outcomes* et de l'impact de trois projets. Dans un second temps, les trois projets ainsi que la méthodologie utilisée et les résultats obtenus sont présentés de façon synthétique dans le tableau 1 ci-dessous. Pour de plus amples informations sur les projets et leur évaluation, le lecteur est invité à lire les rapports projets publiés en annexe.

Le projet du Rwanda n'est pas repris car le CRED n'a pas été en mesure de collecter légalement toutes les données de seconde main nécessaires à l'analyse quantitative, ceci malgré le respect de la démarche exigée par la Ministre de la santé : l'analyse quantitative via un protocole de recherche scientifique avait été présentée en vue d'obtenir l'aval du *National Health Research Committee*.

3.1 Evaluation d'un projet d'appui à l'enseignement technique et professionnel en RDC (AETP2)

La méthodologie retenue pour cette évaluation est **une étude de cas avec contrefactuel dans la ville de Kisangani**. Cette étude se base donc sur une analyse qualitative rigoureuse des observations du terrain et des informations collectées par le biais d'entretiens individuels et de groupes menés dans des écoles bénéficiaires et des écoles témoins, ainsi qu'auprès de personnes clefs liées au secteur et/ou au projet en Belgique et en RDC. Des entretiens approfondis ont été menés avec des directions d'école, des professeurs, des élèves et des anciens élèves diplômés. Etant donné l'absence de donnée quantitative exploitable pour ce projet et vu que la mise en œuvre d'un dispositif de collecte de données de première main s'est révélée peu pertinente suite aux constats de la mission exploratoire, l'application de méthodes quantitatives a dû être abandonnée.

L'exercice démontre que **le projet AETP2 n'a eu aucun effet tangible sur les bénéficiaires du projet dans la ville de Kisangani**. La qualité de l'enseignement n'a pas été améliorée ; les effectifs n'ont pas augmenté ; la capacité d'autofinancement des établissements scolaires n'existe pas ; et l'employabilité des jeunes diplômés n'a pas évoluée. Cet échec s'explique en partie par l'utilisation très partielle des « outputs », mais surtout par l'environnement très défavorable du secteur de l'enseignement technique et professionnel dans le pays. D'une part, il existe des défaillances d'infrastructure qui caractérisent ce pays depuis de nombreuses décennies, encore amplifiées par les situations de conflits violents. D'autre part, l'absence de systèmes incitatifs adéquats et même l'existence d'incitants pervers nuisent à l'efficacité de toute action visant à améliorer la qualité de l'enseignement. A cet égard, le rapport est riche d'informations extrêmement précises et détaillées.

Par ailleurs, des éléments nous amènent à penser que ces constats ne sont pas propres à la ville de Kisangani, car le contexte institutionnel dans lequel le projet s'insère est celui du pays tout entier. De plus, la logique d'intervention et le mode de suivi-évaluation de la CTB sont identiques dans toutes les provinces. Ceci implique que même si l'atteinte des « outputs » est meilleure dans certaines écoles, l'impact du projet AETP2 n'est certainement pas plus important ailleurs (sauf peut-être quelques exceptions liées principalement à la personnalité des acteurs).

Une extrapolation des constats de l'étude de cas est proposée, pour porter un jugement argumenté sur la faible atteinte des *outcomes* dans l'ensemble des écoles bénéficiaires et l'absence d'impact global de ce projet. Notons que même si l'ensemble des *outcomes* avaient été atteint, l'impact à l'échelle nationale ou même provinciale n'aurait pas été significatif, car les projets AETP 1&2 ne touchaient que 5% des écoles professionnelles et techniques du pays.

Malheureusement, l'enseignement principal qui doit en être tiré est que dans les conditions mises en lumière, le programme d'appui à l'enseignement technique et professionnel n'aurait pas dû avoir lieu ou aurait dû être conçu différemment. Il est en effet impossible d'avoir une action d'appui à l'enseignement efficace lorsque l'Etat lui-même n'a pas mis en place une politique adéquate appuyée par des moyens budgétaires appropriés. **En RDC, loin d'être porteuse, la politique d'enseignement est pénalisante.** D'abord, l'Etat ne paie qu'une partie du corps enseignant (les enseignants dits « mécanisés »). Ensuite, les salaires versés sont nettement insuffisants pour assurer un niveau de vie décent, ce qui oblige les parents d'élèves à compléter les salaires des professeurs. Le coût de l'école pour les ménages est dès lors très élevé car il faut ajouter à ces contributions de « motivation » une série d'autres frais qui comprennent le financement non seulement des livres et des uniformes mais aussi des consommables requis pour le fonctionnement des équipements nécessaires à la formation. Enfin, en plus de réserver des montants dérisoires au financement de l'enseignement, l'Etat congolais impose une taxe sur les contributions des parents aux salaires des enseignants. A ces difficultés, s'ajoutent tous les problèmes causés par l'infrastructure défaillante du pays, en particulier l'absence fréquente d'électricité qui compromet la mise en fonctionnement des équipements techniques.

Ces constats traduisent une connaissance superficielle du secteur et **une gestion plus orientée sur la réalisation des « outputs » que sur l'atteinte des *outcomes* et de l'impact** poursuivis.

Bien que l'appréciation contenue dans le rapport se base sur un type de méthodologie simple (une étude de cas avec contrefactuel dans la ville de Kisangani), **elle a le mérite de mettre en lumière des aspects-clé de l'échec.** Ces facteurs inhibant tout effet d'une telle intervention sont suffisamment bien étayés pour convaincre que la conclusion resterait sans doute valide si une méthodologie plus sophistiquée avait pu être menée dans des conditions satisfaisantes. L'utilisation d'un contrefactuel n'ajoute pas grand-chose à la solidité du diagnostic posé. Mais ceci n'aurait pas pu être découvert avant la réalisation du travail de terrain. Le contrefactuel a permis de vérifier les critères de sélection des établissements scolaires mis en place par le projet, et de trianguler certaines informations, ainsi que de comprendre encore mieux le contexte et les enjeux aux différents niveaux.

3.2 Evaluation d'un projet d'amélioration et de renforcement des points d'eau au Sénégal (PARPEBA)

En l'absence de données de base exploitables, un design quasi-expérimental a pu être mis en œuvre afin de collecter des **données de première main auprès d'un groupe cible** (bénéficiaires d'un accès à l'eau du forage) **et d'un groupe témoin** (non-bénéficiaires directs, car ils ne disposent pas d'accès à l'eau du forage au sein de leur village). Les analyses effectuées tantôt par l'application de méthodes quantitatives (double différence), tantôt par une argumentation qualitative, vont dans de nombreux détails et contiennent de riches enseignements pour le donateur et les responsables locaux. Cette richesse d'information tient beaucoup au fait que l'analyse s'est fondée sur des enquêtes de ménages et à la présence relativement longue des évaluateurs sur le terrain. Cela a permis d'approfondir de nombreux aspects des comportements et attitudes des utilisateurs ruraux de l'eau.

Bien que des effets positifs de l'accès à l'eau du forage aient été démontrés, **la qualité de l'eau tant au niveau bactériologique que chimique reste un réel problème**. En effet, plus de 2/3 des sites construits ou réhabilités par PARPEBA présentent des teneurs en fluor et en sel nocives pour l'organisme. De plus, les sources de contamination de l'eau entre le point d'approvisionnement et de consommation sont nombreuses. La population n'est pas suffisamment sensibilisée à ces problématiques de qualité de l'eau. Un tel constat appelle des mesures urgentes de correction au risque de causer des problèmes graves de santé aux populations rurales dans un avenir proche. Hormis l'énorme problème de qualité de l'eau, le projet a globalement eu **un effet positif sur les bénéficiaires et dans une moindre mesure sur les non-bénéficiaires**. En effet, ces derniers consomment eux aussi l'eau du forage, bien qu'ils n'y aient pas accès directement dans leur village.

L'étude démontre que l'utilisation de l'eau du forage a permis **un gain de temps et d'énergie significatif, spécialement pour ceux ayant accès à un robinet privé** (financé sur fonds propres). Le temps et l'énergie gagnés ne semblent pas avoir d'effet sur la consommation d'eau du ménage. Cependant, ces gains semblent encourager le démarrage de petites activités commerciales, spécialement en saison sèche, ainsi que la participation dans des groupements et associations locales. Ces deux derniers effets peuvent évidemment être liés. Une enquête plus approfondie serait nécessaire pour éclairer cette question dans la mesure où l'enquête ne repère pas d'accroissement de revenus chez les ménages bénéficiaires.

Un autre effet des nouveaux forages ressortant de l'enquête microéconomique réalisée est le **gain de bien-être et l'amélioration de la cohésion sociale** (diminution des conflits entre les femmes, principales responsables de la collecte de l'eau). Notons que ce résultat est mesuré sur base d'opinions émises et non sur des éléments d'observation factuelle, il convient donc de rester prudent car il pourrait y avoir un biais de réponse. On ne peut en effet exclure la possibilité que des utilisateurs veuillent donner une image positive des effets du projet dans une enquête manifestement destinée à en évaluer l'impact. L'effet sur l'éducation, à savoir une diminution des écarts dans la fréquentation scolaire entre genres, mérite également d'être approfondi car énormément de facteurs exogènes au projet peuvent expliquer l'évolution observée. Un des grands mérites de l'enquête de ménages effectuée au Sénégal est qu'elle met en lumière un certain nombre *d'effets indirects non-attendus* de l'action de coopération entreprise.

Il est impossible de quantifier la contribution du PARPEBA à l'augmentation de la couverture d'accès à l'eau dans la région du bassin arachidier, car la situation de référence est confuse et le mode de calcul à effectuer n'est pas clair dans la littérature du projet, et du secteur. Cependant, par une estimation très grossière, il est montré que **le PARPEBA a contribué à offrir un accès à l'eau amélioré à minimum 15.5% de la population de la région touchée**. Malheureusement, ce pourcentage est à diviser par 3, si l'on ne considère que les forages offrant une eau de bonne composition chimique.

3.3 Evaluation d'impact d'un projet de petite et moyenne hydraulique au Maroc (PMH)

En absence de données de base exploitables, un **design quasi-expérimental a également pu être mis en œuvre afin de comparer un groupe cible** (agriculteurs bénéficiant des réhabilitations des systèmes traditionnels d'irrigation de crue) **avec un groupe témoin** (agriculteurs n'ayant pas bénéficié de réhabilitation). Les données collectées par le biais d'enquêtes de ménages ont été soumises à deux méthodes statistiques : des tests de double différence grâce à la reconstitution de certaines des données de base en faisant appel à la mémoire des répondants ; et l'utilisation de la technique de l'appariement (« *Propensity Score Matching* ») réalisé majoritairement sur des données post-intervention.

L'analyse montre que grâce aux opérations de réhabilitation du système d'irrigation de crues, le projet a eu un **effet positif sur les superficies cultivées en période de**

3. Principales conclusions sur l'atteinte des « outcomes » et de l'impact des 3 projets

stress hydrique, alors qu'aucune amélioration n'est observée en période de crue. Par ailleurs, le projet ne semble pas avoir encouragé la diversification des cultures ni des changements de pratiques culturales. Il semblerait que, si le projet n'a pas résulté dans une diminution de la variation interannuelle des rendements agricoles (d'où la persistance des stratégies de diversification basées sur la migration), il a néanmoins réussi à **élever le revenu agricole moyen des agriculteurs**. Par rapport aux agriculteurs non-bénéficiaires, les bénéficiaires du projet tirent une plus grande partie de leurs revenus de l'activité agricole non seulement en période de sécheresse mais aussi en période de crue. Cette dernière conclusion est à prendre avec précaution car elle se base sur une appréciation du revenu par les répondants eux-mêmes. Il faudrait des observations de rendements (non disponibles dans les données collectées) s'étalant sur plusieurs années pour pouvoir conclure avec certitude que le projet a permis d'élever le revenu attendu des agriculteurs concernés.

En outre, le projet a encouragé l'extension spontanée des réseaux d'irrigation surtout dans les périmètres en amont. De plus, il a potentiellement permis de recharger la nappe phréatique.

Manifestement, ce projet mérite d'être suivi et évalué au cours de plusieurs années successives car, en dépit de ses faiblesses, il présente un grand intérêt. En effet, même s'il faut rester prudent dans l'interprétation des résultats, il semblerait que les agriculteurs dépendant de ce type d'irrigation ont vu leur niveau de vie s'améliorer de manière perceptible grâce au projet. Etant donné que le projet PMH ne touche pas moins de 70% des terres en irrigation de crue de la province de Tiznit, il existe **un impact globalement positif**.

Tableau 1 : Présentation synthétique des 3 rapports d'évaluation

	Projet d'appui à l'enseignement technique et professionnel en RDC (AETP2)	Projet d'amélioration et de renforcement des points d'eau dans le bassin arachidier au Sénégal (PARPEBA)	Projet de petites et moyennes hydrauliques dans la région de Tiznit au Maroc (PMH)
DONNEES DU PROJET			
Années et budget	2005-2008 / 4 100 000 euros	2003-2008 / 14 976 444 euros	2004-2007 / 960 000 euros
Zone géographique	8 provinces	Le bassin arachidier (3 régions)	La plaine de Tiznit (3 communes)
Politique sectorielle	Inexistante	Existante. L'accès à l'eau prime sur la qualité de l'eau.	Existante. La participation des bénéficiaires est refusée.
Activités principales	<ul style="list-style-type: none"> - Réhabilitation de 27 ateliers pour 6 filières porteuses dans 21 écoles - Livraison d'équipements techniques et didactiques - Formation de professeurs - Sensibilisation à l'ETP - Encourager des AGR au sein des écoles - Création d'unités formation/emploi au sein des écoles 	<ul style="list-style-type: none"> - Réalisation de 32 forages multi-villages - Réhabilitation de 14 forages - Création et renforcement des capacités des ASUFOR - Techniques de potabilisation de l'eau (1 station d'osmose inverse + 1village test avec os calciné) - Volet assainissement (14 édicules publics + formation de 2 femmes relais hygiène par ASUFOR) 	<ul style="list-style-type: none"> - Réhabilitation de 13 barrages et 27 km de canaux - Protection de 22 km de canaux (plantations sur les talus, protection mécanique des ouvrages, etc) - Distribution et plantation de 52 000 arbres fruitiers - Formation à l'arboriculture fruitière et greffe sur caroubiers - Formation à la gestion administrative et financière des AUEA
Bénéficiaires	Ecoles, Professeurs, Elèves	Population rurale	Agriculteurs
Outcomes attendus	<ul style="list-style-type: none"> (1) l'attractivité pour l'ETP dans les filières soutenues des écoles appuyées ; (2) l'amélioration de la qualité de l'ETP dans les écoles soutenues ; (3) la création de revenus et la capacité d'autofinancement des écoles AETP2 ; (4) l'employabilité des jeunes diplômés des filières/écoles soutenues. 	<ul style="list-style-type: none"> (1) Accès à une source d'eau améliorée ; (2) Accès durable à l'eau ; (3) Accès à une quantité d'eau suffisante à un prix accessible ; (4) Accès à une eau de qualité satisfaisante → Gain de temps et d'énergie, amélioration de la situation socio-économique et de la santé 	<ul style="list-style-type: none"> (1) Augmentation des superficies irrigables (2) Augmentation des superficies cultivées (3) Diversification des cultures (4) Intensification de l'activité agricole : plus de récoltes par saison agricole (5) Meilleure gestion des ressources en eau et du système d'irrigation → Gain de temps et développement de nouvelles activités
Impacts attendus	<ul style="list-style-type: none"> - L'amélioration de l'ETP en RDC (Participation à l'atteinte de l'OMD 3 -éducation pour tous-) 	<ul style="list-style-type: none"> - Pérennisation des systèmes d'alimentation en eau - Participation à l'OMD 7 (taux de couverture d'accès à l'eau potable) - Amélioration des conditions de vie des populations rurales de la zone 	<ul style="list-style-type: none"> - Revenus des agriculteurs assurés et stabilisés - Diminution de l'exode rural - Amélioration des conditions de vie et de travail des populations rurales

3. Principales conclusions sur l'atteinte des « outcomes » et de l'impact des 3 projets

	Projet d'appui à l'enseignement technique et professionnel en RDC (AETP2)	Projet d'amélioration et de renforcement des points d'eau dans le bassin arachidier au Sénégal (PARPEBA)	Projet de petites et moyennes hydrauliques dans la région de Tiznit au Maroc (PMH)
APPROCHE METHODOLOGIQUE			
Données pré-projet exploitables ? (« baseline »)	NON	En théorie, oui, mais PAS accessibles en pratique.	NON
Données collectées par le système de SE exploitables ?	NON, système de SE orienté sur les activités mises en œuvre (« outputs »).	NON, système de SE orienté sur le suivi opérationnel (« outputs »).	NON, système de SE orienté sur le suivi opérationnel (« outputs »).
Mission exploratoire	OUI à Kinshasa : test des outils d'enquêtes à Kinshasa (écoles cibles AETP1 et témoins), organisation logistique, entretien avec personnes clés DGD/CTB/ministères	OUI dans le bassin arachidier : focus group avec ASUFOR, organisation logistique pour l'enquête de ménage, identification du contrefactuel, test du questionnaire village et ménage	OUI dans la plaine de Tiznit : identification d'un contrefactuel, création des bases de sondage et des échantillons, organisation logistique pour l'enquête de ménage, test du questionnaire ménage
Outils de collecte de données	Entretiens avec personnes clés à Bruxelles et à Kinshasa Etude de cas avec contrefactuel dans la ville de Kisangani : 2 écoles cibles / 2 écoles témoins ; 4 filières (2 par écoles). - Entretien avec 4 directions d'école - 2 Focus group avec professeurs - 3 Focus group avec élèves de 6 ^{ème} - 1 Focus group avec diplômés - Récolte de données statistiques de seconde main	Entretiens avec personnes clés à Bruxelles, à Dakar et dans les trois régions touchées par le projet Enquête de ménages + focus groups avec contrefactuel dans la région de Kaolack, nouveaux forages : 2 forages offrant une eau de bonne qualité chimique, 2 autres présentant une trop forte teneur en fluor et sel ; - 4 Focus group ASUFOR - 30 Focus group autorités villageoise - 217 Enquêtes de ménages	Entretiens avec personnes clés à Bruxelles, à Rabah et à Tiznit. Enquête de ménages avec contrefactuel dans la plaine de Tiznit (6 communes concernées : 3 pour le groupe cible et 3 pour le contrefactuel) : 9 périmètres ayant bénéficié du projet et 5 villages dont le système d'irrigation reste traditionnel - 231 enquêtes : 129 dans le groupe cible et 102 dans le contrefactuel
Outils d'analyse	- Reconstruction de la logique d'intervention - Analyses statistiques de données de 2 nd e main - Comparaisons d'ordre qualitatif des différents groupes cibles et témoins	- Reconstruction de la logique d'intervention - Test de différence de moyennes et d'indépendance de facteurs sur données 1 ^{ère} main (niveau ménage et individus) - Tests en double différence	- Reconstruction de la logique d'intervention - Test de différence de moyennes et d'indépendance de facteurs sur données 1 ^{ère} main - Estimations économétrique : modèle de Différence-en-Différence et technique du « <i>Propensity Score Matching</i> »

3. Principales conclusions sur l'atteinte des « outcomes » et de l'impact des 3 projets

	Projet d'appui à l'enseignement technique et professionnel en RDC (AETP2)	Projet d'amélioration et de renforcement des points d'eau dans le bassin arachidier au Sénégal (PARPEBA)	Projet de petites et moyennes hydrauliques dans la région de Tiznit au Maroc (PMH)
CONSTATS PRINCIPAUX SUR L'ATTEINTE DES OUTCOMES ET DE L'IMPACT			
Outcomes	<ul style="list-style-type: none"> - peu d'effets significatifs sur les bénéficiaires car pas ou peu d'utilisation des outputs. - prolongation de la capacité à enseigner l'ensemble des cours pratiques au sein des écoles soutenues grâce au renouvellement du matériel ancien et la livraison de petits outils - amélioration éphémère de la qualité de la formation grâce à l'utilisation de consommables livrés dans les écoles soutenues (aujourd'hui inexistant) - extrapolation possible des conclusions issues de l'étude de cas à l'ensemble des zones touchées 	<ul style="list-style-type: none"> - Le PARPEBA a contribué à fournir un <u>accès amélioré et durable à l'eau</u> à la population des villages raccordés mais également, dans une moindre mesure, à la population des villages voisins non-raccordés - Les observations de terrains semblent indiquer que le PARPEBA n'a pas eu l'effet d'accroître les quantités d'eau consommée par jour et par personne. - Gains de temps et d'énergie significatifs (surtout si accès à un robinet privé) - Réallocation très probable de ce temps et énergie pour des AGR, du bien-être et de la cohésion sociale (p-ê aussi dans l'éducation des filles) - Haute disposition à payer l'eau du robinet - 2/3 des forages offrent une eau de mauvaise qualité chimique, dangers à long terme sur la santé - volets assainissement et potabilisation de l'eau peu convaincants - autres effets positifs et négatifs inattendus (cf. rapport) 	<ul style="list-style-type: none"> - Le projet PMH n'a pas eu d'effet sur la capacité à irriguer en période de crue. En revanche, le projet a eu un <u>effet positif en période de faible crue</u> (plus grandes superficies irriguées parmi les bénéficiaires). - Les observations de terrain indiquent que le projet PMH ne semble pas avoir eu un effet sur la capacité des ménages bénéficiaires à mettre en culture une proportion plus importante de leurs terres irriguées. - Le projet ne semble pas avoir eu un effet positif sur le maintien ou l'abandon de l'activité agricole. - Les analyses ne permettent pas de conclure que le projet ait eu l'effet d'accroître le nombre de récoltes par saison agricole pour les bénéficiaires. - Les bénéficiaires du projet PMH tirent une plus grande proportion de leurs revenus de l'activité agricole aussi bien en période de crue qu'en période de sécheresse. - Pas de preuves que le projet ait contribué à une diversification des cultures
Impact	<ul style="list-style-type: none"> - pas d'impact au niveau de l'amélioration de la qualité de l'ETP en RDC - à l'échelle nationale, les projets AETP 1&2 touchent 5% des écoles, donc impossible d'avoir un impact significatif au niveau national même si il y avait un impact au niveau des écoles appuyées - l'ensemble des interventions belges dans le secteur a contribué au moins indirectement aux prémices de l'amélioration de l'ETP (différentes réformes sont en cours) 	<ul style="list-style-type: none"> - Le PARPEBA contribue significativement et de manière durable à l'amélioration des conditions de vie d'une part non négligeable de la population des trois régions concernées - Mais il y a un grand risque sur la santé de la population car 2/3 des forages offrent une eau non potable. - Difficultés d'apprécier sa contribution à l'OMD7 car situation de référence confuse, ainsi que la mesure des indicateurs à considérer 	<ul style="list-style-type: none"> - La contribution du projet PMH est significative dans la mesure où le projet touche plus de 70% des terres en irrigation de crue de la Province de Tiznit. - Pas de preuves quant à la réduction de l'exode rurale. - Pas de preuves quant à une amélioration notable des conditions de vie des bénéficiaires et à un niveau plus global des conditions sociales de la région - Deux impacts inattendus : une poursuite de l'extension des réseaux (essentiellement en amont) ; et le projet aurait un effet positif sur la recharge des nappes phréatiques.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d'outcomes et d'impact

Cette section présente d'abord les enseignements de type méthodologique sur le processus d'évaluation dans son ensemble. Elle expose ensuite des enseignements sur la conception, la mise en œuvre et le suivi-évaluation des interventions. En effet, l'évaluation d'impact soulève des questions sur la manière dont se déroule le cycle du projet.

4.1 Enseignements sur le processus d'évaluation

L'application de méthodes mixtes sur base de la logique d'intervention est une solution adéquate pour apprécier et expliquer les *outcomes* et impacts d'un projet.

La « **theory based evaluation** » est donc une approche globale pouvant utiliser tant les méthodes quantitatives que qualitatives pour apprécier les *outcomes* et l'impact d'un projet.

Le **processus évaluatif combinant des méthodes qualitatives et quantitatives dans l'analyse du cadre logique** a permis :

- **De mesurer et démontrer l'attribution de certains *outcomes* des projets évalués.** En effet, à cause d'une absence de données de base (« baseline ») et de la difficulté à utiliser des données de « *recall* » fiables pour reconstruire des indicateurs préalable au projet, seule une minorité des *outcomes* attribuables aux projets du Sénégal et du Maroc¹⁶ a pu être analysée par l'application de méthodes quantitatives.
- **D'apprécier et d'argumenter l'attribution au projet de l'ensemble des autres *outcomes* non traitables quantitativement par une analyse qualitative rigoureuse.**
- **De porter un jugement argumenté et rigoureux sur l'atteinte de l'impact des projets évalués ;**
- De mettre en évidence les mécanismes et les éléments du contexte expliquant les effets ou non-effets constatés. Les évaluations apportent un éclairage sur les facteurs inhibant ou favorisant l'atteinte des *outcomes* et de l'impact. **Comprendre le 'pourquoi' et le 'comment' des effets ou de l'absence d'effet d'un projet** est un aspect clef d'une évaluation d'impact.

Aucune des évaluations ex post réalisées n'était un cas idéal pour appliquer les méthodes quantitatives : aucune données de base disponibles, des contrefactuels imparfaits, des échantillons relativement petits, une contrainte de temps serrée.¹⁷

¹⁶ Pour l'évaluation des projets au Sénégal et au Maroc, des enquêtes de ménages ont été réalisées, ce qui a permis l'utilisation de techniques quantitatives.

¹⁷ Voir « *Conducting impact evaluation under budget, time and data constraints* » (IEG, 2013), pour des explications sur les décisions méthodologiques à prendre dans de telles situations.

Cependant, ce n'est pas pour autant qu'une évaluation quantitative ait été abandonnée. In fine, elle apporte un éclairage différent sur l'atteinte des « outcomes », même si c'est pour une minorité et permet d'accroître la rigueur des analyses qualitatives. De plus, elle met en évidence que seule une analyse quantitative est en mesure de démontrer l'attribution des effets au projet. Cette question de l'attribution est centrale dans l'évaluation d'impact.

L'exercice mené montre que pour apprécier les *outcomes* et l'impact d'un projet, la question n'est pas de choisir l'une ou l'autre méthode mais de **combiner de façon adéquate les méthodes qualitatives avec les méthodes quantitatives**. En effet, les méthodes quantitatives lorsqu'on peut les appliquer permettent d'objectiver la mesure et l'attribution des effets d'une intervention sur les bénéficiaires. En d'autres termes, elles permettent de démontrer empiriquement l'atteinte des « outcomes ». Mais elles ne se prononcent pas sur l'atteinte de l'impact (effet du projet à un niveau global, cf. définition supra). Elles servent tout au plus comme argument à l'analyse qualitative de l'impact, par le biais d'extrapolation ou de généralisation de certains constats démontrés empiriquement. Les méthodes qualitatives posent, quant à elles, un jugement argumenté de l'atteinte des *outcomes* non traitables quantitativement et sur l'atteinte de l'impact (effet du projet à un niveau global). Finalement, des explications sur l'atteinte ou l'échec des résultats escomptés des projets ont été proposées grâce à une connaissance approfondie du contexte et du projet acquise par la mise en place du processus évaluatif combinant les deux méthodes (cf. infra).

Bien que la littérature et ces expériences montrent la pertinence de combiner des méthodes quantitatives avec des méthodes qualitatives pour apprécier les effets d'un projet, cela ne signifie pas que cela soit facile et pertinent pour tout type d'intervention.

Alors que les méthodes qualitatives sont quasi toujours applicables, l'application des méthodes quantitatives nécessite des conditions supplémentaires : des données fiables en quantité suffisante pour des indicateurs pertinentes mesurés avant le projet et après le projet auprès d'un groupe de bénéficiaires et d'un contrefactuel crédible. **En plus des données fiables et pertinentes en quantité suffisante, l'application de méthodes mixtes nécessite du temps**. Ces deux conditions sont loin des réalités du monde de l'évaluation, où *'il faut aller vite'*, et loin des réalités des projets de la coopération bilatérale belge, puisque *'très peu de données exploitables sont disponibles'*.

Cependant, se priver d'utiliser des méthodes quantitatives, c'est se priver d'une mesure objective des effets sur les bénéficiaires et d'une preuve scientifique de leur attribution au projet. C'est également se priver d'un grand nombre d'observations qualitatives utilisables pour construire une argumentation rigoureuse de l'atteinte des *outcomes* non quantifiables et de l'impact. Il convient donc de juger au préalable de la faisabilité d'appliquer les deux types de méthodes.

L'exercice mené ne cherche pas à comparer les méthodes d'analyse, comme c'est souvent le cas dans la littérature. En effet, d'une part, les méthodes quantitatives ne sont pas réalisables sans dimension qualitative. D'autre part, les méthodes qualitatives ne sont pas substituables aux méthodes quantitatives dans la mesure où elles ne répondent pas aux mêmes questions : elles sont incapables de mesurer et de démontrer empiriquement les « outcomes ». Cependant, il est essentiel de comprendre leurs spécificités et leur complémentarité pour les utiliser au mieux dans le cadre d'une évaluation d'impact. Les avantages tirés de leurs complémentarités et spécificités pour évaluer les effets des projets sont détaillées ci-dessous.

4.1.1 Un dispositif d'évaluation quantitative enrichit l'analyse qualitative

La décision d'utiliser des méthodes quantitatives a enrichi l'évaluation qualitative sur trois plans.

- Tout d'abord, grâce une analyse avec contrefactuel, à l'utilisation de statistiques et de modèles économétriques, l'application de méthodes quantitatives permet de **mesurer de façon précise les effets (ou au moins une partie des effets) d'un projet sur les bénéficiaires et d'en démontrer scientifiquement leurs liens de causalité au projet**, ce qu'une analyse qualitative n'est pas en mesure de faire. Une telle démonstration empirique des effets d'un projet est plus convaincante et rigoureuse qu'une argumentation factuelle de l'atteinte des *outcomes* (analyse qualitative).
- Ensuite, bien que les analyses quantitatives n'aient pas pour objectif de se prononcer sur l'atteinte de l'impact global d'un projet, certains constats chiffrés sur les *outcomes* peuvent être extrapolés et parfois généralisés. Cela signifie que **certains résultats quantitatifs peuvent aider à apprécier l'impact global**, ou tout au moins aider à apporter des arguments forts en faveur ou non de l'atteinte de l'impact d'un projet.
- Finalement, via le dispositif de collecte de données de première main mis en place pour l'application des méthodes quantitatives, **les données qualitatives collectées sont plus nombreuses et de meilleure qualité** que si elles avaient été collectées dans la cadre d'une analyse qualitative classique. En effet, un dispositif quantitatif impose que ces données soient **collectées de façon aléatoire directement sur un grand nombre de bénéficiaires et non-bénéficiaires**, et ce parfois à différents niveaux pour un même projet (par exemple au niveau des ménages, des villages, des directions d'école, des professeurs, des gestionnaire de forage, etc.).

Par ailleurs, le mode de collecte sur le terrain a aussi son importance : les **missions exploratoires** sont riches en collecte de données qualitatives ; les **enquêtes de ménages** regorgent d'informations qualitatives exploitables statistiquement ; et la **présence relativement longue sur le terrain** permet de relever énormément d'observations d'intérêt¹⁸.

Le fait de disposer d'une plus grande variété de données qualitatives fiables augmente significativement les possibilités pour apprécier l'impact global. De plus, toutes ces données qualitatives collectées grâce à la mise en place d'un processus d'évaluation quantitatif aident à la compréhension des mécanismes et des éléments du contexte inhibant ou favorisant l'atteinte des *outcomes* et de l'impact.

Un dispositif de collecte de données pour l'application uniquement de méthodes qualitatives ne permet pas de fournir cette quantité et ce niveau de détail dans les données qualitatives.

Les trois évaluations réalisées sont de bons exemples pour illustrer **comment le dispositif quantitatif a permis de prendre conscience qu'il est possible de faire des analyses qualitatives d'une plus grande rigueur**. Le dispositif quantitatif mis en œuvre et le temps passé sur le terrain ont réellement permis de fournir une argumentation rigoureuse sur l'atteinte des résultats.

Le dispositif évaluatif s'est mis en place de la manière suivante :

¹⁸ En moyenne 4 semaines pour réaliser une enquête de ménage, dans le cadre des évaluations réalisées. Cela peut être significativement plus long selon l'ampleur du projet et les objectifs poursuivis par l'étude.

- (i) **Mission exploratoire pour préparer la collecte d'information.** Afin de concevoir les questionnaires pour l'enquête de ménages et les guides d'entretiens pour les focus group, l'équipe n'a pas pu se contenter de lire les rapports existants sur les projets. Elle a dû aller sur le terrain afin de vérifier certains faits relatés dans les documents et se rendre compte du contexte actuel par le biais de rencontres avec les bénéficiaires et les autorités locales.

Ces missions exploratoires se sont toutes révélées très utiles.

Au **Maroc**, elle a permis de réaliser que la liste des villages bénéficiaires trouvée dans les rapports comportait beaucoup d'erreurs et qu'il n'existait aucune liste des paysans bénéficiaires au sein des villages sélectionnés par le projet. Ces listes sont bien évidemment importantes si l'on veut se rendre dans les villages pour y mener une enquête. Ces constats ont également mis en évidence une série de dysfonctionnements du projet.

Au **Sénégal**, la mission exploratoire a permis de mieux comprendre certains faits affirmés dans les rapports qui jettent un éclairage sur l'atteinte de l'impact. Par exemple dans les rapports, les bornes privées apparaissent comme un « output » du projet. Or, c'est un « outcome » car elles sont financées par les utilisateurs eux-mêmes grâce au réseau de distribution construit par le projet. Ensuite, toujours selon les rapports, certaines bornes fontaines ne seraient plus utilisées par manque d'entretien alors qu'en réalité elles ne sont plus utilisées parce qu'elles sont devenues inutiles du fait que la majorité des utilisateurs ont aujourd'hui accès à l'eau dans leur concession.

En **RDC**, la mission exploratoire à Kinshasa a permis de comprendre une série d'incitants pervers dans les systèmes institutionnels inhibant tout effet possible du projet. Il a néanmoins été décidé de poursuivre l'exercice d'évaluation en se rendant en province afin de trianguler les premiers constats.

Au **Rwanda**, une telle mission n'a pas été programmée. Elle aurait vraisemblablement permis de se rendre plus vite compte de la complexité de mener l'évaluation du projet dans le contexte politique actuel.

- (ii) **Collecte de données de première main.** Il existe un niveau de détails important dans les questionnaires de ménages. Ces outils de collecte permettent de poser de nombreuses questions d'ordre qualitatif (appréciation et questions ouvertes). Les réponses à ces questions ont pu être codifiées, encodées et donc analysées via des méthodes statistiques et/ou utilisées pour formuler une argumentation rigoureuse.

Les questionnaires utilisés dans l'évaluation des projets au **Sénégal** et au **Maroc** illustrent le niveau de détail et le type de données qualitatives collectées et analysées (voir dans les annexes des rapports pays).

- (iii) **Temps passé sur le terrain** pour la collecte de données. Par une présence relativement longue sur le terrain, les évaluateurs ont le temps de réaliser des entretiens plus long, d'observer, et d'obtenir des informations ad hoc non contenues dans les guides d'entretien ou les questionnaires. L'ensemble de ces éléments aident à l'interprétation de certains constats.

Au **Sénégal**, une simple observation de la localisation des bornes publiques dans les villages visités a permis de comprendre que les utilisateurs d'une borne publique parcouraient la même distance qu'avant en allant au puits. En effet, le puits et la borne fontaine ont été quasi

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

systématiquement localisés au même endroit. Aucun gain de distance à parcourir ne pouvait donc être espéré contrairement à ce qu'affirmaient plusieurs documents. Certains effets inattendus ont également été découverts grâce à la présence de l'évaluateur sur le terrain. Par exemple, l'équipe accueillie dans un poste de santé a pu observer que l'accès à l'eau via un robinet augmentait la présence des fonctionnaires (professeurs, sages-femmes) dans les villages.

Au **Maroc**, la présence sur le terrain a permis à l'évaluateur de se rendre compte de certaines manipulations dans les réponses données par les répondants, mais également de critiquer plus sévèrement la pertinence du contrefactuel.

En **RDC**, le niveau de développement des infrastructures observé dans les écoles a permis de comprendre aisément pourquoi du matériel moderne n'avait aucune chance d'être utilisé. Par ailleurs, le cas de la RDC est particulier car aucune méthode quantitative n'a été appliquée, étant donné la révision des choix méthodologiques suite à la mission exploratoire (annulation des enquêtes de ménages). Cependant, le processus d'évaluation présente certains aspects quantitatifs tels que le choix aléatoire des écoles, l'utilisation d'un contrefactuel, la formulation des hypothèses sous-jacentes à chaque « outcome » et les propositions de tests pour apprécier les *outcomes* moyennant l'existence de données (et plus fondamentalement moyennant l'existence d'un impact).

4.1.2 Une bonne analyse qualitative est nécessaire à la mise en œuvre rigoureuse des méthodes quantitatives

Les méthodes quantitatives enrichissent véritablement l'évaluation des effets d'un projet comme cela a été démontré ci-dessus. Toutefois, pour être efficace, l'application de telles méthodes nécessite une mise en œuvre rigoureuse qui soulève plusieurs défis (cf. supra pour une explication détaillée) : (i) trouver un bon contrefactuel en minimisant le problème de biais de sélection et de contagion ; (ii) utiliser des données de qualité en quantité suffisante ; (iii) appliquer correctement les techniques statistiques et économétriques ; et (iv) interpréter les résultats des analyses quantitatives de façon correcte et en cohérence avec les réalités du terrain.

Une bonne analyse qualitative est nécessaire pour faire face à ces nombreux défis qui risquent de ruiner une analyse quantitative, spécialement en situation d'évaluation ex post. A chacune des phases du dispositif quantitatif d'évaluation, l'analyse qualitative est utilisée pour justifier les choix méthodologiques et pour interpréter les résultats des analyses quantitatives.

Les phases du dispositif évaluatif quantitatif ex post sont exposées ci-dessous en illustrant tant que possible la rigueur requise et l'utilité de l'analyse qualitative pour sa mise en œuvre. Chacune de ces étapes est cruciale et peut être une source d'erreurs qui conditionne la qualité des évaluations réalisées avec l'aide des méthodes quantitatives.

(1) La phase préparatoire

- Rien ne peut commencer sans une **bonne compréhension de la logique d'intervention du projet, de la thématique du projet et du contexte tant du pays, que de l'intervention ou encore du contexte au moment de l'évaluation.** Tout ceci relève des méthodes qualitatives.
- Pour l'application des méthodes quantitatives, il s'agit ensuite **d'identifier les outcomes mesurables et le type de données nécessaires** pour les quantifier et en démontrer leur attribution avec le projet.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

- Cela implique, à ce stade, de **définir avec précision les indicateurs pertinents** pour réaliser les analyses d'*outcomes* souhaitées. Notons qu'il est indispensable de décliner avec précision la définition des indicateurs et de préciser les sources d'informations utilisées pour les construire. Sans des données précises, il devient difficile d'interpréter les résultats.

L'exemple qui suit illustre l'importance de la précision dans la définition d'un indicateur. Un indicateur du projet PARPEBA au **Sénégal** est le « taux de couverture d'accès à l'eau potable ». ¹⁹ Mais de quel accès parle-t-on : borne privée ou publique, à quelle distance minimale de l'habitation ? Comment définit-on une eau potable : composition chimique et/ou bactériologique ? De quelle eau parle-t-on : l'eau à la sortie du château d'eau ou à la sortie du robinet ou encore au point de consommation ?

L'exemple suivant montre l'importance des sources des données (et leur fiabilité). Pour le projet en **RDC**, il existe plusieurs façons de mesurer le taux de scolarité : via les registres de présence (les taux de présence seront vraisemblablement différents si l'on prend les présences le matin ou en fin de journée, en saison des pluies ou en saison sèche) ou via une enquête de ménage ou encore via les registres au niveau provincial (avec des risques de manipulation des données). Que l'on en considère l'une ou l'autre source pour construire cet indicateur, les analyses seront vraisemblablement différentes.

Angus Deaton insiste sur cela dans sa contribution à la Conférence sur l'évaluation d'impact organisée par l'AFD et EUDN en 2010 (« *measure for measure* ») : « *different data lead to different conclusions* ».

Les personnes en charge d'une évaluation doivent donc se défendre d'avoir les données appropriées pour leur analyse en donnant des gages de qualité. Une façon est de présenter avec détail la méthodologie suivie, et les biais potentiels pouvant résulter de l'utilisation de certaines données.

- La phase préparatoire est également le moment de **concevoir des outils de collecte de données** (questionnaires, guide d'entretien) et de **choisir un contrefactuel crédible**.
- Ceci n'est possible qu'à partir d'une bonne connaissance du terrain, acquise entre autres grâce à une **mission exploratoire**. Suite à cette mission et sur base d'observations qualitatives fondées, les choix méthodologiques sont argumentés dans l'objectif de surmonter divers biais potentiels identifiés (biais de sélection, contagion, pertinence du contrefactuel...). Les outils d'enquêtes sont également améliorés.
- C'est pendant la phase préparatoire, qu'il faut rentrer en contact avec le **partenaire local** et identifier un **accompagnateur de terrain**.

Le premier est nécessaire pour l'organisation logistique de la mise en œuvre de collecte de données de première main. Il aide à rassembler une équipe d'enquêteurs fiables et compétents, à organiser les transports, il donne des conseils concernant le logement de l'équipe, etc. Idéalement, ce partenaire doit avoir des connaissances en évaluation quantitative mais aussi être conscient des réalités du terrain. Des centres de recherche ou des universités peuvent être de bons partenaires.

Le second est une personne proche du projet sur le terrain sans qu'il y ait de conflit d'intérêt. Il est nécessaire que son rôle soit clairement défini et que sa présence n'ait pas de conséquence sur l'attitude des répondants. Être accompagné sur le terrain permet de s'introduire plus facilement auprès des

¹⁹ Par ailleurs, notons en passant que les définitions des OMD sont extrêmement imprécises. Il est donc vraiment difficile de se prononcer sur la contribution d'un projet à l'atteinte des OMD.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact populations locales et de gagner leur confiance en dépassant certaines barrières culturelles.

Sans de tels soutiens, la mise en place d'un dispositif rigoureux de collecte de données de première main est quasi impossible.

- Cette mission exploratoire est aussi l'occasion **d'obtenir l'aval des autorités nationales du pays** pour mener ce genre d'étude.

Au **Rwanda**, la Ministre de la santé a manifesté de fortes réticences face à l'évaluation du projet sélectionné peut être par peur que les constats puissent avoir des retombées au niveau politique. Bien que l'équipe et le SES aient tenté de gagner la confiance des autorités et d'obtenir leur aval pour la réalisation de la collecte de données de seconde main, l'ensemble des autorisations n'a pas été obtenu. L'évaluation a donc dû être arrêtée²⁰.

(2) La phase d'enquête

- Il s'agit tout d'abord de **tester les outils de collecte** de données pour garantir une bonne formulation et une bonne compréhension des questions, surtout dans la langue locale.
- Il faut aussi **vérifier les compétences et la fiabilité des enquêteurs**.
- Il convient sans cesse de **s'adapter aux réalités du terrain sans perdre de vue la rigueur nécessaire** pour être en mesure d'appliquer des méthodes quantitatives aux données collectées. Les aléas du terrain sont nombreux et divers, un pneu crevé, des conditions climatiques difficiles, l'absence ou l'indisponibilité des répondants, des voies de communication impraticables, etc. Il est fréquent de devoir revoir en partie la taille ou le type d'échantillon.

Il faut être conscient que de réduire de façon significative l'échantillon ou modifier la zone géographique couverte peut remettre en question la faisabilité de l'utilisation de méthodes quantitatives (utilisation d'outils économétriques et statistiques) et donc remettre en question la crédibilité des analyses.

Il s'agit donc de s'assurer qu'on n'ait pas introduit de biais importants nuisibles à la qualité des estimations des « outcomes ». Pour minimiser ces biais, plusieurs actions sont utiles : collecter un maximum d'informations pouvant **démontrer la pertinence du contrefactuel** (observations qualitatives et indicateurs quantitatifs), ainsi que **la pertinence et fiabilité des données collectées**. Il est important de prendre le temps de gagner la confiance des répondants afin d'éviter des réponses manipulées menant à des surestimations ou sous-estimations des effets. Il est indispensable de relire les questionnaires en fin de journée et demander des explications aux enquêteurs en cas d'incohérences ou de non réponse. Le cas échéant, il faut renvoyer l'enquêteur auprès des répondants pour leur poser des questions supplémentaires.

- Finalement, l'étape qui clôture la phase de terrain consiste à obtenir une base de données exploitable. Cela implique de créer un **masque de saisie** sur un programme informatique (CSPPro par exemple) et trouver une équipe pour **encoder les données collectées via les questionnaires**. Ceci doit se faire sous la supervision d'une personne expérimentée. En effet, avoir de bonnes données sur papier n'est pas une garantie d'avoir des données bien encodées, et des données mal encodées donnent lieu à des analyses erronées.

²⁰ Malheureusement, tout ceci a été découvert lors de la mission de terrain, car une mission exploratoire n'a pas pu être planifiée compte tenu des contraintes budgétaires de l'exercice.

(3) **La phase d'analyse**

- Tout d'abord, il faut **nettoyer les données** pour vérifier la cohérence des informations encodées avec celles collectées sur papier.
- Il faut ensuite choisir des **méthodes quantitatives adaptées** au type de données disponibles et aux biais potentiels identifiés (problème de sélection, fiabilité et pertinence des données « baseline », validité du contrefactuel, etc).
- Il ne suffit pas d'identifier les méthodes, il faut veiller à **les utiliser correctement** (respect des hypothèses). STATA, un logiciel statistique et économétrique, a été utilisé pour l'analyse quantitative.
- Ensuite, il faut s'assurer de la **robustesse des résultats** (résultats identiques à travers différents modèles et avec différentes spécifications)
- Il convient enfin **d'interpréter les résultats à la lumière des observations du terrain**. Rester critique face aux chiffres avancés par les méthodes est essentiel surtout s'ils contredisent des intuitions de terrain. Dans ce cas, il faut vérifier qu'il n'y a pas une erreur de manipulation technique. L'analyse qualitative a toute son importance dans cette dernière étape.

4.1.3 L'analyse qualitative indispensable pour comprendre les effets d'un projet (« outcomes »/impact) et pour apprécier l'impact

La logique d'action et le contexte sont deux éléments clés pour comprendre une intervention et son impact. En plus d'être utiles pour justifier la mise en place d'un dispositif quantitatif, départager les choix méthodologiques et interpréter les constats sur les « outcomes », l'analyse de la logique d'intervention et du contexte sont indispensables pour porter un jugement argumenté sur l'atteinte de l'effet à un niveau global d'une intervention (impact). En effet, les méthodes quantitatives se limitent à l'appréciation des « outcomes ». Cependant, les constats empiriques sur les *outcomes* peuvent servir d'arguments pour l'appréciation qualitative de l'impact (cf. supra).

Mettre en évidence la **chaîne de causalité** entre les ressources mobilisées et les objectifs à atteindre permet tout d'abord d'avoir un regard critique sur la faisabilité de l'intervention. Les actions menées sont-elles susceptibles de conduire aux objectifs ? Ces objectifs sont-ils réalistes aux vues du contexte et des ressources mobilisées ? Quels sont les risques et les hypothèses sous-jacentes ? Toutes ces questions permettent de comprendre l'intervention de sa conception à l'impact.

La connaissance du **contexte** de l'intervention est déterminante pour comprendre l'impact d'une intervention. La situation socio-économique, politique et culturelle au moment de l'intervention, les relations diplomatiques entre les pays, les relations entre bailleurs, la compréhension des enjeux de leadership à différents niveaux de pouvoir et la perspective historique de l'intervention²¹ sont autant d'éléments contextuels pouvant avoir une influence sur l'atteinte ou non de l'impact d'une intervention.

Pour les trois évaluations, **ces aspects se sont révélés essentiels pour l'analyse et ont permis de montrer les insuffisances, tant dans la conception que dans la mise en œuvre ou le suivi des projets, qui empêchent ou réduisent les impacts attendus** (cf. infra).

²¹ Par exemples : existaient-ils d'autres interventions de ce type ? Du même bailleur ? D'autres bailleurs ? Ont-elles été bien ou mal accueillies par la population ? D'autres intervention de ce type sont-elles prévues ?

4.1.4 Un dispositif qui permet de considérer l'ensemble des acteurs

Une combinaison des deux approches méthodologiques permet **d'approfondir les entretiens avec les bénéficiaires grâce à l'approche quantitative et d'apprécier avec plus de précision le rôle de l'ensemble des acteurs gravitant autour d'une intervention grâce à l'approche qualitative.**

En effet, la mise en œuvre de méthodes qualitatives nécessite de collecter des informations auprès de la CTB, de la DGD, également auprès du partenaire (ministère en charge du secteur, autorités locales, etc.), et de quelques bénéficiaires. Ces informations sont très utiles, entre autres, pour contextualiser l'intervention. Le dispositif évaluatif quantitatif met, quant à lui, la priorité sur les bénéficiaires choisis en grand nombre et de façon aléatoire. La combinaison de ces deux méthodes permet donc de mener des entretiens plus approfondis avec les différents acteurs en jeu pour la réussite d'une intervention. Cela présente l'avantage de tenir compte des spécificités de chaque acteur et de les sensibiliser à l'évaluation d'impact.

4.1.5 Faisabilité et pertinence d'une évaluation quantitative ex post

Étant donné l'exigence et la rigueur nécessaires pour réaliser des évaluations utilisant des méthodes quantitatives, ainsi que le temps et les coûts que cela représente, la question de la faisabilité d'une évaluation quantitative a été soulevée pour les 4 projets à évaluer. Une bonne analyse qualitative de la logique d'action et du contexte au moment du projet et lors de l'évaluation, complétée par une mission exploratoire a permis de se prononcer sur la pertinence de mettre en place un dispositif évaluatif quantitatif.

Selon Naudet et al. (2012), la méthode expérimentale est adaptée à des projets qu'ils caractérisent de « projet tunnel », c'est-à-dire des interventions caractérisées par un nombre limité d'intrants homogènes, une chaîne de causalité claire et testée, une participation large et des effets rapides. Son analyse n'est que peu pertinente dans le cas des quatre évaluations menées, car ce sont des évaluations ex post de conception quasi expérimentale. Malgré des conditions qui étaient loin d'être optimales pour réaliser des analyses quantitatives, la méthodologie initiale proposée pour les quatre projets comportait systématiquement une dimension quantitative. Cependant, in fine, seuls deux évaluations (Sénégal, Maroc) sur les quatre font usage de méthodes quantitatives.

Ci-dessous, sont exposées les situations des quatre projets en termes de faisabilité d'une analyse quantitative. Ensuite, des conclusions sont tirées de ces expériences. Finalement est abordée la question de la pertinence de mener des évaluations quantitatives alors que d'autres en ont mené sur des thématiques similaires.

1) Faisabilité d'une évaluation quantitative dans les 4 projets évalués

a) L'appui à l'enseignement technique et professionnel en RDC :

- **Faits** : ce projet a une portée quasi nationale, avec une certaine mise en œuvre séquentielle car le projet antérieur, relativement similaire, touchait d'autres provinces. Il n'a pas été possible d'utiliser des données de seconde main pour deux raisons : il n'existe pas de données récentes de bonne qualité collectées au niveau des provinces, et surtout la probabilité de voir des effets du projet au niveau d'une province était très faible vu le très petit nombre d'établissements scolaires touchés par le projet dans chaque province. Il n'existait pas non plus de données disponibles au niveau des projets (ni de la phase 1, ni de la phase 2). En outre, il était impossible de se rendre dans toutes les provinces pour collecter des informations. Un dispositif de collecte par téléphone ou courrier

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

électronique n'était pas non plus envisageable dans le contexte de sous-développement des infrastructures dans ce pays.

- Proposition initiale : la note méthodologique initiale proposait une étude de cas avec contrefactuel dans une province choisie sur base de critères objectifs avec des enquêtes auprès de différents types de bénéficiaires (écoles, professeurs, élèves, anciens élèves diplômés).
- Constats après la mission exploratoire : cette mission a permis de réaliser que la probabilité de ne pas trouver d'effet du projet sur les bénéficiaires était très élevée. La mise en place d'un dispositif de collecte de première main via des enquêtes individuelles a donc été remise en question.
- Analyses réalisées : l'approche méthodologique est restée la même avec le remplacement des enquêtes individuelles par des « focus group » auprès des mêmes groupes de bénéficiaires et d'un contrefactuel. Des entretiens approfondis ont été menés avec les acteurs du secteur, mais également avec des directions d'écoles, des professeurs, des élèves et des anciens élèves diplômés. Des données de seconde main au niveau des établissements scolaires ont également été collectées. L'ensemble des données collectées a permis de comparer des indicateurs pertinents pour chaque « outcome » entre les écoles cibles et les écoles témoins pour 4 filières d'intérêt. Ces comparaisons couplées d'une analyse du contexte institutionnel a permis une analyse qualitative rigoureuse sur la non-atteinte des effets escomptés du projet tant au niveau des bénéficiaires, qu'à un niveau global.

b) Appui à la construction d'un laboratoire de référence en santé publique au Rwanda :

- Faits : ce projet a une portée nationale, avec une implémentation séquentielle du contrôle qualité. Il existe des données de seconde main de relativement bonne qualité au niveau des districts et des laboratoires.
- Proposition initiale : il a été proposé d'utiliser ces données de seconde main, en utilisant la mise en œuvre séquentielle pour créer un contrefactuel valide. Des entretiens approfondis avec un échantillon de chefs de laboratoires, de centres de santé, et d'hôpitaux étaient également prévus.
- Constats pendant la mission terrain (il n'y a pas eu de mission exploratoire): les données sont en effet disponibles au moins pour étudier un aspect du projet (le contrôle des analyses de la goutte épaisse). Cependant l'équipe n'ayant jamais reçu l'autorisation officielle pour les collecter officiellement et donc de les analyser, l'évaluation n'a pas pu être menée à son terme (même si la partie qualitative a pu se dérouler pendant la mission).

c) Projet d'adduction d'eau potable et amélioration sanitaire au Sénégal :

- Faits : ce projet est concentré sur 3 régions du bassin arachidier sans mise en œuvre séquentielle. Le projet comporte différents volets et différentes dimensions : accès à l'eau via la réhabilitation ou la construction de forages, techniques de potabilisation de l'eau pour tenter de diminuer les teneurs en fluor dans deux sites pilotes, constructions d'ouvrages sanitaires et sensibilisation à l'hygiène. Des diagnostics participatifs existent. Ils contiennent des informations d'ordre quantitatif et qualitatif renseignant sur la situation préalable à l'intervention dans certains sites. Cependant, les données brutes ne sont plus disponibles, elles ne sont donc pas exploitables pour l'analyse d'impact ex post. Aucun contrefactuel n'était prédéfini dans le projet.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

- Proposition initiale : la note méthodologique proposait la réalisation d'une enquête de ménages avec contrefactuel dans une des régions touchées par le projet en se concentrant sur les effets de l'accès à l'eau via les nouveaux forages. Le volet assainissement, ainsi que le volet potabilisation de l'eau n'ont pas fait l'objet d'une conception méthodologique quantitative par manque de temps et de moyens. En plus d'enquêtes de ménages, des « focus group » au niveau des villages et des gestionnaires des forages ont été proposés. La reconstitution de données de base était prévue grâce à des données « recall » et via les informations disponibles dans les diagnostics participatifs.
- Constats après la mission exploratoire : il était possible de trouver un contrefactuel crédible dans la zone d'enquête même imparfait avec un échantillon petit. Des biais de surestimation des effets étaient possibles car le groupe témoin était moins desservi en eau que le groupe cible, et cela, déjà avant l'intervention. De plus, un effet de contagion existait : des ménages vivant dans des villages non raccordés à un forage, s'alimentaient également en eau d'un forage. Il était possible de collecter des données sur la situation pré-intervention (« recall ») mais pas pour tous les indicateurs d'« outcomes ». Par conséquent, cela limitait l'utilisation de méthodes quantitatives. Il a également été observé que la participation des bénéficiaires était quasi unanime. De ce fait le biais de sélection était peu probable et par conséquent la petite taille de l'échantillon posait moins de problème. Le partenaire local identifié était d'un bon niveau et avait déjà rassemblé une équipe d'enquêteurs. L'accompagnateur de terrain trouvé convenait bien pour la mission. L'accueil dans les villages était chaleureux, les répondants se rendaient disponibles facilement.
- Analyses quantitatives réalisées : les données collectées ont permis d'utiliser les méthodes quantitatives sur deux *outcomes* (effet du projet sur la distance à parcourir pour chercher de l'eau et sur le temps alloué à cette tâche). D'autres *outcome* sont fait l'objet d'analyses statistiques sans pour autant être en mesure de démontrer leur attribution au projet (effet sur l'éducation ou le bien-être par exemple). Les données collectées ont par ailleurs permis d'enrichir l'analyse qualitative à plusieurs niveaux.

d) Projet de petite et moyenne hydraulique au Maroc :

- Faits : ce projet est concentré sur la plaine de Tiznit sur trois communes. Il porte principalement sur la réhabilitation de systèmes traditionnels d'irrigation de crues dans des villages situés le long de la rivière principale. Il comporte également d'autres volets : la reforestation et le renforcement des capacités tant au niveau des techniques culturelles que de la gestion des ouvrages réhabilités. Le projet n'a pas de données de base, ni de contrefactuel prédéfini.
- Proposition initiale : il avait été prévu de réaliser une enquête de ménages avec contrefactuel couvrant les 3 communes touchées par le projet et 3 autres communes « témoins » situées sur un affluent de la rivière principale.
- Constats après la mission exploratoire : il n'existe pas de données, même au niveau des villages, des bénéficiaires du projet. Créer l'échantillon n'était pas facile. Le contact avec les autorités locales s'avérait difficile. Elles ont tenté d'influencer certains choix méthodologiques et les réponses de certaines personnes interrogées. Le partenaire local n'était pas expérimenté dans ce type d'enquêtes individuelles. Il a aussi été constaté qu'il serait impossible de collecter

des données quantitatives sur le rendement des terres avec le temps et les budgets disponibles.

- Analyses quantitatives réalisées : les méthodes quantitatives ont pu être appliquées avec succès pour deux *outcomes* (les superficies irriguées en période de grande crue et de faible crue ; la part de revenus agricoles dans le revenu total). Il convient toutefois de rester prudent quant à l'interprétation de ces résultats, qui pourraient être sur estimés, étant donné les doutes soulevés à propos de la crédibilité du contrefactuel sur certains aspects. De plus, l'appariement ayant été fait sur des données ex post et certains résultats n'étant pas robustes à travers diverses spécifications, le rapport reste prudent quant à l'interprétation des résultats trouvés.

2) Conclusion sur la faisabilité d'analyses quantitatives dans l'évaluation

Des expériences menées, il ressort **quatre faits** influençant la faisabilité d'une analyse quantitative des effets d'une intervention.

- a) L'essentiel est d'avoir un **contrefactuel crédible** pour être en mesure de réaliser des évaluations quantitatives des « outcomes », dans le sens « *attribution analysis* » (3ie).

L'utilisation d'outils quantitatifs est fortement remise en cause

- lorsque le projet a une portée nationale sans mise en œuvre séquentielle de l'intervention (connu sous le nom anglais de « *pipeline process* ») ; ou
 - lorsque les problèmes de contagion et de biais de sélection ne peuvent être éliminés ou tout au moins minimisés par une argumentation qualitative rigoureuse et/ou par des techniques économétriques solides.
- b) Il est nécessaire d'avoir ou d'être en mesure de reconstruire des données de base. La « **baseline** » est un autre élément important pour mener des « *attribution analysis* ». Sans baseline, il n'est pas impossible d'utiliser certains outils quantitatifs, mais leur application est plus difficilement justifiables (cf. supra, appariement sur des données ex post).
 - c) La faisabilité d'analyses quantitatives de l'impact dépend aussi de **l'accord des autorités locales**. En effet, des analyses quantitatives rigoureuses ont l'avantage de fournir des preuves de l'existence d'effets (ou d'absence d'effets) liés au projet. Ces constats scientifiquement prouvés peuvent donc avoir des répercussions politiques plus conséquentes car moins facilement réfutables qu'un jugement argumenté sur l'atteinte des effets d'un projet tel que proposé par une analyse qualitative.
 - d) La « **value for money** » de la mise en œuvre d'un processus d'évaluation quantitative. Si une analyse qualitative rigoureuse permet de montrer que la probabilité d'observer un effet est très faible, il est alors inutile de collecter des données de première main. En effet, les analyses quantitatives de ces données n'apporteraient rien de plus. De plus, il faudrait un très grand échantillon pour justifier que l'absence d'effet n'est pas due à la taille de l'échantillon. Le contraire n'est pas vrai : si une analyse qualitative montre une probabilité raisonnable d'observer un effet, il est alors utile de mettre en œuvre un dispositif quantitatif d'analyse. En effet, ce dernier permettra de se prononcer sur l'ampleur des effets et d'en démontrer empiriquement les liens de causalité avec le projet. C'est plus rigoureux et convaincant que n'importe quelle analyse qualitative.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

Par exemple, en **RDC**, il a été décidé d'abandonner la réalisation d'enquêtes de ménage après avoir démontré que le projet ATEP2 dans le contexte spécifique de l'enseignement technique et professionnel dans le pays avait une haute probabilité de ne pas atteindre les *outcomes* escomptés et donc l'impact attendu. Les efforts à déployer pour réaliser une analyse quantitative n'étaient ainsi plus justifiés.

Notons que la question de « value for money » touche un aspect bien plus large des évaluations d'impact. L'analyse coûts/bénéfices ne sera jamais positive si les bénéfices d'une évaluation quantitative ne sont pas perçus. Cette perception renvoie à l'utilisation des constats d'une évaluation. En d'autres termes, quel besoin aurait-on de faire des évaluations si c'est pour ne pas en utiliser les constats ? Le coût onéreux de mettre en œuvre un dispositif d'évaluation quantitative ne peut jamais être estimé acceptable si la mesure des effets sur les bénéficiaires et la preuve de leur attribution au projet ne présentent pas d'intérêt pour les décideurs et/ou praticiens du développement.

Trois remarques méritent encore attention. Elles concernent la portée des analyses quantitatives.

- a) Les méthodes quantitatives peuvent être appliquées pour apprécier **un aspect d'un projet** (un volet en particulier ou même une dimension d'un volet), mais plus difficilement pour évaluer l'ensemble des activités d'une intervention. On préférera des effets facilement identifiables et mesurables. Les autres aspects d'un projet et les *outcomes* difficilement mesurables peuvent alors être analysés qualitativement.
- b) Même si le projet a une portée nationale, ou touche plusieurs régions, on peut réaliser une analyse quantitative des *outcomes* sur **une zone géographique restreinte**. L'analyse dans une province, commune ou région apporte déjà de nombreux enseignements. Ensuite, une bonne connaissance des réalités dans les autres régions permet parfois d'extrapoler ou même de généraliser certains constats à l'ensemble des zones touchées par le projet.
- c) Des **données de seconde main de qualité** suffisent parfois pour réaliser des analyses quantitatives, spécialement si elles sont accompagnées d'une analyse qualitative rigoureuse. L'idéal est que ces données aient été générées par le processus de suivi-évaluation du projet (cf. point suivant).

L'exercice mené dans le cadre de ces quatre évaluations ex post d'impact montre que même dans une situation sous-optimale, il est possible et pertinent de mettre en œuvre un dispositif d'évaluation par des méthodes quantitatives si certaines conditions sont remplies.²² Ces constats ont mené à l'élaboration du tableau 2 ci-dessous.

Ce tableau résume les conditions à remplir pour assurer la **faisabilité d'évaluations quantitatives rigoureuses de conception quasi-expérimentale** (par opposition avec la conception expérimentale du « *randomized control trial* » peu réaliste dans le monde de la coopération bilatérale belge) **dans une situation ex ante et ex post**.

Selon que l'on planifie l'évaluation des *outcomes* et de l'impact au moment de la conception du projet (évaluation ex ante) ou à la fin d'un projet (évaluation ex post), certaines conditions doivent être remplies pour assurer la faisabilité et la pertinence d'un dispositif quantitatif d'évaluation d'un projet. L'objectif poursuivi est de quantifier les

²² Pour une discussion approfondie sur comment mener des évaluations sous contraintes temporelles et budgétaire serrées, avec un accès limité à des données, il est conseillé de lire l'article suivant : « Conducting quality impact evaluations under budget, time and data constraints » -Impact evaluation group (IEG) de la Banque Mondiale-.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

changements observés sur les bénéficiaires suite à l'intervention et d'en démontrer empiriquement les liens de causalité.

Dans une conception quasi-expérimentale, les conditions minimales et optimales influencent le niveau de qualité des analyses quantitatives possibles.

La qualité d'une analyse quantitative dépend, en plus de la maîtrise des outils, des éléments suivants :

- De la **pertinence du contrefactuel** identifié. Il convient de prouver qu'il est valable par des analyses statistiques mais également par un jugement argumenté basé sur l'analyse qualitative du contexte et observations du terrain.
 - o Si un projet connaît des effets de contagion trop importants sur la population non bénéficiaire, l'application de méthodes quantitatives sera sérieusement remise en cause.
 - o La mise en œuvre séquentielle d'une intervention (« *pipeline process* ») est une option qui est satisfaisante sur le plan méthodologique et qui n'est pas trop contraignante pour la mise en œuvre du projet.
- De la **quantité des données** collectées.
- De la **fiabilité et pertinence des données** à analyser. Des données « *recall* » sont toujours jugées moins fiables que des données factuelles au moment des enquêtes. Il convient également de justifier la qualité des sources des données de seconde main utilisées.
- De la **faculté à minimiser ou au moins à identifier les éléments pouvant biaiser les résultats** (biais de contagion et de sélection principalement), via des modèles économétriques et des arguments tirés des observations et entretiens réalisés sur le terrain.

Deux éléments méritent d'être soulevés :

- La « **baseline** », même bien faite, est une condition nécessaire mais pas suffisante pour être en mesure d'appliquer des méthodes quantitatives d'évaluation.
- La décision de concevoir une **évaluation d'impact ex ante** offre plus d'opportunités pour être en mesure de réaliser des évaluations quantitatives. Par ailleurs, une conception ex ante peut amener à modifier la mise en œuvre d'un projet. Par exemple, on peut privilégier une mise en œuvre séquentielle pour créer un contrefactuel pertinent (« *pipeline process* »).

Les conditions nécessaires pour la réussite d'une évaluation quantitative sont également ajoutées au tableau. Sans le respect de ces conditions, il est inutile de procéder à la mise en œuvre d'un dispositif quantitatif d'évaluation. Finalement, les trois remarques concernant la portée des évaluations quantitatives sont résumées dans la dernière ligne du tableau sous l'intitulé 'alternatives'. Ces alternatives permettent de limiter l'ampleur de l'évaluation quantitative et donc inévitablement son coût, mais a des conséquences sur le nombre d'effets attribuables qu'on sera en mesure d'analyser.

Insistons sur le fait que même si les conditions optimales sont remplies, il est **impossible de faire l'impasse sur une analyse qualitative rigoureuse**. Elle est indispensable pour argumenter certains choix méthodologiques et pour d'interpréter les résultats quantitatifs. **Il ne suffit de démontrer ce qui fonctionne ou ce qui ne fonctionne pas, il faut encore expliquer pourquoi et comment on arrive à de tels constats.**

Tableau 2 : Conditions pour la réalisation d'évaluations quantitatives de conception quasi-expérimentale, ex ante ou ex post à l'intervention.

Conception quasi-expérimentale	Conception ex ante de l'évaluation	Conception ex post de l'évaluation
Conditions nécessaires	<ul style="list-style-type: none"> - Aval des autorités locales - Collaboration du partenaire - Incitants à mettre en œuvre un dispositif de suivi-évaluation axé sur les résultats (ressources humaines, budget, compétences) 	<ul style="list-style-type: none"> - Aval des autorités locales ; - Probabilité relativement élevée de trouver un impact - Disposer de suffisamment de temps et d'un budget suffisamment élevé - Prévoir une mission exploratoire
Conditions minimales	<ul style="list-style-type: none"> - Identifier un contrefactuel crédible (sans effet trop important de contagion) - Données « baseline » fiables et pertinentes en quantité suffisante sur un échantillon de bénéficiaires - Dispositif de suivi-évaluation permettant une gestion axée sur les résultats (outcome/impact) → données de monitoring - Données ex post fiables et pertinentes en quantité suffisante sur les bénéficiaires et le contrefactuel - Situation de base (« baseline ») du contrefactuel reconstruite sur base de données « recall » et/ou données de 2nde main de qualité. 	<ul style="list-style-type: none"> - Identifier un contrefactuel crédible (sans effet trop important de contagion) - Données ex post fiables et pertinentes en quantité suffisante sur les bénéficiaires et le contrefactuel - Situation de base (« baseline ») des bénéficiaires et du contrefactuel reconstruite sur base de données « recall » et/ou données de 2nde main.
Conditions optimales	<ul style="list-style-type: none"> - Identification d'un contrefactuel crédible (sans effet trop important de contagion), grâce par exemple à la mise en œuvre séquentielle de l'intervention (« pipeline process ») - Données « baseline » fiables et pertinentes en quantité suffisante sur un échantillon de bénéficiaires - Identifier au moment de la mise en œuvre du projet, une source de données de 2nde main de qualité pour reconstruire la « baseline » du contrefactuel - Dispositif de suivi-évaluation permettant une gestion axée sur les résultats (outcome/impact) → données de monitoring - Données ex post fiables et pertinentes en quantité suffisante sur les bénéficiaires et le contrefactuel - Collecter des données « recall » pour compléter la « baseline » du contrefactuel 	<ul style="list-style-type: none"> - Identifier un contrefactuel crédible (sans effet trop important de contagion), le cas échéant utiliser la mise en œuvre séquentielle du projet - Données ex post fiables et pertinentes en quantité suffisante sur les bénéficiaires et le contrefactuel - Situation « baseline » des bénéficiaires et du contrefactuel reconstruite sur base de données « recall » et/ou données de 2nde main (via le dispositif de suivi-évaluation qui a collecté les indicateurs pertinents –nécessité d'avoir accès aux données brutes-).
Alternatives	<ul style="list-style-type: none"> - Concentrer l'évaluation quantitative sur une dimension du projet ou sur une zone géographique - Selon les objectifs poursuivis et le type d'intervention, réduire le nombre d'enquêtes individuelles pour se concentrer sur un niveau plus élevé de bénéficiaires (villages, communes, écoles, hôpitaux, centres de santé, etc.). 	

3) Raisons pour mener des évaluations quantitatives bien que des évidences existent

Un des objectifs de l'évaluation est de capitaliser des expériences, c'est-à-dire de réaliser un grand nombre d'études démontrant ce qui fonctionne et ce qui ne fonctionne pas en matière d'aide au développement. Certains pourraient donc être tentés de ne pas chercher à évaluer l'impact de leur intervention, mais de calquer leur intervention sur un projet qui a réussi. Dans un tel cas, il suffirait ensuite de concentrer les efforts sur le suivi et non sur l'évaluation.

Il est certes intéressant et important de s'inspirer des conclusions d'évaluations quantitatives existantes pour concevoir un projet. Cependant, partir de l'hypothèse que le projet mis en place produira les effets désirés car il s'inspire d'un projet similaire déjà évalué de façon positive est risqué, et cela pour plusieurs raisons.

- a) **La validité externe limitée des « *Randomized control trial* ».** Les évaluations quantitatives les plus à même à fournir des preuves du succès d'une intervention sont les RCT. Cependant, ces dernières sont souvent menées à petite échelle et en grande partie mises en œuvre par des chercheurs académiques (conditions optimales recherchées pour appliquer des méthodes quantitatives). Il n'est donc pas à exclure que les résultats soient relativement différents, une fois le projet élaboré et mis en œuvre à plus grande échelle par une agence de développement avec les partenaires locaux et sans soutien académique. (White, 2013, 3ie).
- b) **L'importance du mécanisme d'intervention.** Les RCT ne cherchent pas toujours à évaluer le mécanisme d'intervention, mais tentent plutôt de répondre à une question de recherche touchant à l'une ou l'autre dimension d'un projet. Or le mécanisme d'intervention est déterminant dans l'aide au développement.

De plus en plus d'études évaluent le mécanisme d'intervention.²³ Cependant, la mise en œuvre de tels mécanismes élaborés dans une perspective de recherche n'est toujours pas évidente tant les conditions du monde réel influencent le processus de mise en œuvre.

- c) **L'importance du contexte et de sa rapide évolution.** Une chose est de copier des interventions ayant fait leurs preuves, une autre est de pouvoir reproduire les conditions qui ont influencé son succès. En effet, toutes les évaluations rigoureuses menées insistent sur l'importance des éléments du contexte favorisant ou inhibant l'atteinte des objectifs d'un projet.

Par ailleurs, il est rare de trouver des évaluations portant à la fois sur les thématiques et la zone géographique d'intérêt. En effet, le contexte au sens large est à considérer comme un élément déterminant du succès d'une intervention : les aspects socio-politico-économiques et culturels, ainsi que la dimension historique.

De plus, il faut également tenir compte de la rapide évolution de la société des pays du Sud (développement des télécommunications, chocs climatiques, etc.) et donc de l'évolution des comportements des individus.

²³ Par exemple pour promouvoir l'éducation, il existe plusieurs mécanismes possibles, entre autres un versement d'argent aux parents pour les encourager à envoyer leurs enfants à l'école (« cash transfers ») ou un programme garantissant un repas à l'école (« food for school »), ou encore le paiement des frais scolaires moyennant certaines sanctions si l'enfant a un taux d'absence trop important. Seule une analyse quantitative rigoureuse des « outcomes » permet de départager les effets de ces différents modes d'intervention.

4.2 Enseignements sur la conception, la mise en œuvre et le suivi-évaluation des projets

Les évaluations d'*outcomes* et de l'impact ont mis en évidence des manquements au niveau de la conception, de la mise en œuvre et du suivi-évaluation des projets.

Force est de constater que sur les trois projets évalués, **le bilan de l'atteinte des *outcomes* et de l'impact global n'est pas bon** si l'on se base sur les objectifs spécifiques et globaux attendus pour chacun des projets. Par contre, si l'on se réfère aux objectifs spécifiques et globaux redéfinis suite à la reconstitution du cadre logique par les évaluateurs, le bilan est légèrement meilleur mais loin d'être bon.

Cela indique deux choses. Tout d'abord, il existe des insuffisances au niveau de la conception/formulation des projets : les objectifs poursuivis sont décalés des réalités du terrain et des activités/moyens de l'intervention. Ensuite le dispositif de mise en œuvre et de suivi-évaluation des projets n'est pas suffisamment efficace pour assurer une gestion axée sur les résultats (*outcomes* et impact).

Certaines améliorations ont certainement été faites dans ces domaines depuis lors, car ces constats concernent des projets terminés depuis 2008. Cependant, bien que les pratiques évoluent, il conviendrait de vérifier si celles-ci pallient les manquements identifiés ci-dessous. Il faudrait ensuite vérifier si ces changements se traduisent réellement sur le terrain et dans les résultats observés (*outcomes* et impact).

4.2.1 Des insuffisances dans la conception/formulation des interventions qui réduisent leur impact potentiel

Dans les 4 projets évalués, plusieurs problèmes de conception/formulation ont été relevés.

Le premier problème est que les projets ont tous une **logique d'action faible** se traduisant par :

- Des **objectifs déconnectés des activités et moyens** mis en œuvre.

Les liens entre les activités et les objectifs spécifiques et globaux n'existent pas toujours ou ne sont pas clairement établis dans la logique d'action. Ce qui a pour conséquence que même si toutes les activités sont réalisées de manière adéquate, les effets sur les bénéficiaires peuvent ne pas se produire, et ces activités peuvent également ne pas contribuer à l'atteinte de l'objectif global.

En **RDC**, le matériel a bien été livré, mais la plupart n'est pas utilisé. Au **Sénégal**, les robinets sont connectés, les gens ont donc accès à l'eau mais la qualité n'est pas garantie, alors que l'objectif était l'accès à l'eau potable.

- Des **objectifs irréalistes** dans le contexte local

Les projets ont souvent des objectifs globaux peu réalistes en décalage avec les ressources mobilisées. C'est le cas par exemple du projet au **Congo** où le projet d'appui technique aux écoles visait à améliorer la situation de l'emploi au Congo alors que ce dernier visait à peine 5% des écoles professionnelles, avec quelques écoles bénéficiaires par province.

L'ensemble des projets est caractérisé par un **manque de connaissance du contexte et des problèmes à résoudre**, et ceci alors que 3 projets sur 4 sont en fait la deuxième phase d'un projet antérieur.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

Par exemple, le projet au **Congo** a pour objectif d'augmenter la disponibilité d'une main d'œuvre technique qualifiée alors qu'il n'y a pas d'emplois dans les filières visées et que l'objectif de presque tous les jeunes qui s'inscrivent dans ces filières est de poursuivre leurs études dans l'enseignement supérieur et non d'aller travailler manuellement. Dans ce même projet, la plupart du matériel livré nécessite une alimentation électrique, alors qu'elle fait défaut même à Kinshasa. Au **Maroc**, le projet a pour objectif de diminuer l'exode rural alors que cette région ne peut vivre qu'avec une diversification de revenus tant les crues et donc l'agriculture de crue est aléatoire, même avec un des systèmes d'irrigation réhabilités. Le travail en ville reste donc pour ces populations une alternative intéressante. Au **Sénégal**, le projet a pour objectif de fournir de l'eau potable, alors que la région a de graves problèmes de potabilité des nappes : 2 forages sur 3 présentent de l'eau à trop forte teneur en fluor et sodium. Toujours au Sénégal, une des techniques de potabilisation de l'eau proposée était la technique des os calcinés alors que cette méthode a été démontrée inefficace dans la littérature depuis une quinzaine d'années.

- Des **objectifs peu précis et mal définis** (pas de valeur cible et/ou pas de définition claire)

Les objectifs manquent systématiquement de précision dans leur définition. Peu d'objectifs sont quantifiés, et il n'est pas clairement spécifié comment mesurer les effets sur les bénéficiaires. De plus, étant donné que les situations de référence n'existent pas, même si certains objectifs sont quantifiés, il est impossible d'apprécier une quelconque évolution.

Par exemple, l'objectif du projet au **Sénégal** était d'augmenter de 50% la couverture du taux d'accès à l'eau potable dans la région, comment apprécier l'évolution sans indicateur pré projet ? Que signifie avoir accès à l'eau, chez soi ou à quelle distance maximum de l'habitation ? Que signifie potable, la composition chimique et/ou bactériologique ?

Un autre problème est **la question de l'alignement sur les politiques nationales**. Selon les engagements sur l'efficacité de l'aide (Déclaration de Paris, Accra, Busan), c'est une condition importante, pour ne pas dire nécessaire, pour l'atteinte de l'impact. Mais l'alignement n'est pas un gage de réussite du projet. En effet, certains éléments des politiques nationales peuvent être contraires au succès de l'intervention tel que défini dans les objectifs spécifiques et globaux.

Par exemple, au moment de la conception du projet PARPEBA au **Sénégal**, le gouvernement soutenait une politique d'accès à l'eau au plus grand nombre sans préoccupation de la qualité. Le projet s'est aligné sur cette politique, ce qui l'a amené à construire ou réhabiliter une majorité de forages présentant des teneurs en fluor et en sel nocives pour l'organisme, alors qu'il voulait fournir de l'eau potable. Au **Maroc**, la dimension participative n'était pas inscrite dans les politiques de développement rural qui privilégiaient strictement les réalisations physiques. Le projet PMH-Tiznit s'est inscrit dans ce cadre politique alors qu'il a depuis longtemps été démontré qu'un projet d'irrigation a plus de chance de réussir et de perdurer lorsque les bénéficiaires participent activement à sa conception et à sa mise en œuvre.

Plus fondamentalement, ces constats sur l'alignement soulèvent la question de la **capacité de négociation avec le pays partenaire** lors de la conception d'un projet. Ils soulèvent également la question de la **capacité à renoncer ou à suspendre une intervention** lorsque la Belgique estime que les conditions pour atteindre les *outcomes* escomptés et l'impact attendus ne sont pas réunies.

Une logique d'action faible et un alignement « aveugle » (c'est-à-dire un alignement à des politiques nationales qui vont à l'encontre de la réussite d'un projet) **diminuent fortement les chances de succès d'une intervention.**

Ces faiblesses constatées dans la conception des projets est préoccupante dans la mesure où elles montrent que **l'impact n'est pas une priorité et que le cadre logique est une formalité pour présenter un projet alors qu'il pourrait être un véritable outil de formulation, de gestion axée sur les résultats et d'évaluation.** En effet, un examen simple des cadres logiques aurait permis de conclure à l'impossibilité d'atteindre la plupart des *outcomes* attendus avant le démarrage des projets évalués.

Ces constats peuvent peut-être en partie être expliqués par le mode d'identification et de formulation des projets au sein même de la coopération bilatérale directe. La DGD dialogue politiquement avec les autorités pour fixer les objectifs spécifiques et globaux des interventions lors de la phase d'identification. Ensuite, la CTB formule un projet dans le but d'atteindre ces objectifs. Cela revient à dire que les inputs et outputs sont définis par la CTB, alors que les *outcomes* et l'impact sont eux définis par la DGD. **Même s'il existe un processus de concertation entre ces deux instances, et le pays partenaire, il n'est pas au point, car il laisse passer des projets irréalistes.**

4.2.2 Une mise en œuvre qui privilégie le décaissement et les réalisations physiques

Pour les quatre projets, les « **outputs** » prévus ont été effectivement réalisés. C'est un constat positif qui mérite d'être souligné dans la mesure où les interventions examinées ont été déployées dans des conditions souvent difficiles au niveau du contexte ou de la relation avec les autorités nationales concernées. L'examen des dossiers de suivi-évaluation révèle les efforts qu'il a fallu déployer pour atteindre ces résultats.

Cependant, qu'en est-il des effets de la réalisation de ces « outputs » ? En d'autres termes, ces « outputs » sont-ils utilisés ? Produisent-ils les effets attendus ?

Force est de constater que cette **gestion axée sur les « outputs » donnait priorité au décaissement et aux réalisations physiques au détriment des objectifs globaux et spécifiques** poursuivis.

Sans gestion axée sur les résultats, au sens *outcomes* et impact, les effets d'un projet deviennent des objectifs secondaires. Cela ressort clairement des projets évalués, par exemple au niveau des décisions prises face à des situations imprévues et/ou anticipées. Au **Sénégal**, lorsque le projet a été confronté aux problèmes d'un nombre important de forages présentant des qualités d'eau non satisfaisante, il a été décidé de ne pas tenir compte des normes et de brancher ces forages. Une telle décision déviait pourtant le projet des objectifs poursuivis. Au **Maroc**, étant donné la topographie des lieux, des études techniques plus approfondies auraient dû être menées sur un certain nombre de périmètres. Cela aurait vraisemblablement augmenté les capacités à irriguer en période de crue. En **RDC**, alors que les écoles n'avaient pas réalisé les réhabilitations qu'elles s'étaient engagées à faire, elles ont quand même reçu le matériel prévu.

Le mode de mise en œuvre observé n'est que **le reflet du manque d'incitants à faire autrement.** Si les exécutants sont appréciés sur le niveau de décaissement et de réalisations des « outputs », pourquoi chercheraient-ils à atteindre les *outcomes* escomptés et l'impact attendu ?

4.2.3 Un système de suivi-évaluation insuffisant

Une autre illustration de cette absence de préoccupation des *outcomes* et de l'impact est le peu d'attention portée à ces aspects dans les dispositifs de suivi-évaluation mis en place. Alors que pour tous les projets, des recommandations relativement précises d'actions à mener en vue d'apprécier les effets du projet sont formulées, celles-ci ne sont jamais mises en œuvre.

4. Enseignements de l'utilisation de méthodes mixtes pour l'évaluation d' « outcomes » et d'impact

Les trois évaluations soulignent des déficiences au sein des **systèmes de suivi des projets**. Bien que des indicateurs pertinents pour mesurer les *outcomes* aient été définis dans les documents de présentation des projets, ces derniers sont souvent mal définis, sans valeur initiale et/ou de valeur cible, sans explication des modalités de collecte (comment, quand, où, par qui ?). Finalement ces indicateurs, même en données de base, sont introuvables quelques années après la fin des projets (par exemple au **Sénégal**, les données brutes ayant normalement conduits à la rédaction des diagnostics participatifs sont indisponibles). Ce constat pose le problème de l'accès à des données fiables de suivi pour réaliser des évaluations sur l'atteinte des résultats des projets (*outcomes* et impact). Par ailleurs, de telles données réduiraient fortement le coût de collecte d'informations lors des évaluations d'impact en fin de projet.

Le **système d'évaluation interne** est également remis en cause. En effet, les conclusions émises par les évaluations finales des projets ont presque toutes été contredites ou nuancées par les constats des évaluations ex post réalisées. Par ailleurs, il n'est pas étonnant de constater que les évaluations finales soient fortement focalisées sur la réalisation des « outputs », vu la priorité donnée au décaissement et aux réalisations physiques.

Ces constats soulèvent une fois de plus **la question des incitants à faire autrement**. Sachant qu'un euro investi dans le système de suivi-évaluation n'est pas investi dans d'autres activités, comment encourager le partenaire et les exécutants des projets à consacrer plus de ressources et de moyens au suivi et à l'évaluation alors qu'ils ne subissent aucune conséquence s'ils ne le font pas et/ou qu'ils ne perçoivent pas bien la valeur ajoutée d'une gestion qui tient compte des résultats, pas uniquement de l'atteinte des « outputs » mais également des « outcomes et de l'impact ? De plus, il ne faut pas négliger l'ampleur du travail de mise en œuvre d'une intervention. Il n'est donc pas aisé de mener de front la mise en œuvre d'un projet et celle d'un dispositif rigoureux de suivi-évaluation.

5. Recommandations

5.1 Définir une stratégie globale d'évaluation où l'évaluation d'impact a une place

Définir, au sein de la coopération belge, une stratégie globale d'évaluation où l'évaluation d'impact a une place.

Un constat : une politique ou stratégie globale d'évaluation pour l'ensemble de la coopération belge n'existe pas, alors que l'évaluation et le rapportage sur les résultats sont prévus dans la loi de mars 2013 sur la coopération internationale.²⁴ Par ailleurs, la Belgique est clairement en retard dans le domaine de l'évaluation d'impact si l'on regarde ce qui se pratique au sein des coopérations bilatérales des pays voisins ou au sein des organisations internationales.

Etant donné la complexité des modalités d'interventions de la coopération belge, le nombre et la spécificité des différentes parties prenantes, une stratégie globale d'évaluation est indispensable. Vu qu'une telle politique n'existe pas aujourd'hui, situer l'évaluation d'impact (un sous-ensemble de l'évaluation) reste un exercice forcément incomplet.

Dans la suite du document lorsqu'on parlera d'évaluation d'impact, on se réfèrera à des **évaluations de projet menées ex post** et visant, par une combinaison de méthodes qualitatives et quantitatives, à mesurer les *outcomes* attribuables à l'intervention et à formuler un jugement rigoureux et argumenté sur l'impact d'une opération de développement. Les *outcomes* sont définis comme les effets sur les bénéficiaires, l'impact correspond aux effets de l'intervention à un niveau global. Ces termes correspondent respectivement aux concepts d'objectifs spécifiques et d'objectif global tels que présentés dans la logique d'intervention.

L'évaluation en général, mais aussi l'évaluation d'*outcomes* et de l'impact poursuit 3 objectifs : **(1) rendre compte ; (2) aider à la décision ; et (3) capitaliser des expériences**. Les constats d'une évaluation, et de surcroît d'une évaluation d'impact, peuvent intéresser les différents partenaires de la coopération (le bailleur et les décideurs politiques, le partenaire, les bénéficiaires, les exécutants de l'aide et les contribuables) qui peuvent les utiliser en fonction de leurs besoins. En effet, l'évaluation d'impact consiste à identifier « ce qui fonctionne et ce qui ne fonctionne pas » par une démonstration empirique, mais aussi à comprendre pourquoi et comment un projet atteint ses objectifs ou au contraire ne les atteint pas.

²⁴ « Art. 32. En vue d'atteindre les objectifs de la coopération belge au développement tels que définis au chapitre 2, les **résultats sont évalués** à la lumière des principes énoncés au chapitre 3 et sur la base des critères fixés par le CAD de l'OCDE, à savoir la pertinence, l'efficacité, l'efficience, la viabilité et l'**impact**, ainsi que sur base de la durabilité. Une approche cohérente sera élaborée à cet effet, en vue de permettre **un rapportage des résultats et une gestion axée sur les résultats**. Un système de rapportage uniformisé devra en outre permettre un **suivi systématique des résultats obtenus et non obtenus**. Il sera en outre tenu compte de ces résultats lors de la décision concernant le déroulement futur des actions entreprises ».

5. Recommandations

Les questions sur l'efficacité de l'aide et sur les types de projets à mettre en œuvre pour générer un impact et avoir un effet sur le développement d'un pays sont au centre des débats. Dans ce contexte, il existe une demande pour **rendre compte** de l'atteinte des *outcomes* et de l'impact des interventions d'aide au développement. Il s'agit de rendre compte aux contribuables, mais également aux partenaires et aux bénéficiaires.

- L'évaluation d'impact peut également **aider à la prise de décision**. Elle peut aider à comprendre l'opportunité de répliquer une intervention ou d'éviter de le faire. Les résultats d'une évaluation peuvent donc aider à prendre des décisions stratégiques en matière d'orientation de l'aide dans un secteur, une région, un pays.

Bien que cela soit théoriquement possible et tout à fait souhaitable, en pratique le « timing » n'est souvent pas favorable pour que l'évaluation d'impact joue un tel rôle. A cet égard, l'exemple de la **RDC** est intéressant. Alors que la décision de continuer un projet dans le secteur de l'enseignement technique et professionnel a déjà été prise, l'évaluation a montré que l'appui dans ce domaine n'a eu aucun impact et qu'il existe aujourd'hui une très faible probabilité qu'un tel appui ait un quelconque effet sur les bénéficiaires dans le contexte actuel.

De plus l'évaluation d'impact peut contribuer à **mieux baser les politiques de coopération sur des preuves de leur efficacité** (« *evidence-based policy making* »). Ainsi la position du ministre de la coopération pourrait être renforcée lors de négociations de budget par exemple. Les décideurs politiques doivent toutefois être prêts à accepter le fait que certains constats rigoureusement démontrés puissent remettre en cause la pertinence de certaines actions ou modes d'intervention dans certains secteurs ou dans certains pays. Mais ne cherche-t-on pas à tendre vers plus d'efficacité dans les dépenses publiques pour plus d'effets sur les bénéficiaires ?

- **Capitaliser les expériences** est un aspect particulièrement intéressant de l'évaluation d'impact. C'est un investissement dans la création d'un bien public. L'évaluation d'impact permet d'identifier les projets qui fonctionnent et ceux qui ne fonctionnent pas, mais elle permet aussi de poser un tel jugement sur un mode d'intervention (façon de concevoir/mettre en œuvre un projet) ou sur une dimension particulière d'un projet. Même si l'ambition n'est pas de répliquer l'intervention ailleurs ou à une plus grande échelle par manque de moyens ou de volonté politique, les résultats d'une évaluation menée avec rigueur peuvent être communiqués. Ils peuvent alors servir d'arguments auprès d'autres bailleurs pour les encourager à financer, ou à l'inverse, les pousser à renoncer à ces types d'interventions.

Le projet du **Maroc** se prête bien à un apprentissage car l'irrigation de crue est encore peu connue. De même, le volet potabilisation de l'eau au **Sénégal** aurait pu faire l'objet d'une évaluation dans un but d'apprentissage tant la problématique de la potabilité de l'eau est grande dans ce pays.

Par ailleurs, la capitalisation peut aussi se faire sur le plan formatif ou méthodologique. Comme c'est, entre autres, le cas pour ces quatre évaluations ex post de l'impact.

La promotion de l'évaluation d'impact et son intégration dans la stratégie globale d'évaluation doit se faire en tenant compte de plusieurs éléments décrits ci-dessous.

5.1.1 Clarifier les définitions de l'impact, de l'« outcome » et de l'évaluation d'impact

Un premier élément important est de **clarifier ce que l'on entend par 'évaluation d'impact', et ce que signifient les termes 'impact' et 'outcome'**. Il convient également de formuler ce que l'on cherche à savoir, quels sont **les objectifs poursuivis** lorsqu'on mène une évaluation d'impact.

En effet, un des constats de cet exercice est la multiplicité des définitions et les malentendus générés lorsque l'on parle d'évaluation d'impact. Entre la DGD, la CTB, le SES et les autres acteurs de coopération, il serait utile d'avoir une définition précise et harmonisée de l'impact et de l'« outcome ». Les définitions présentées dans ce rapport pourraient facilement servir de base à cette discussion (cf. supra, section 3).

A noter que ce débat sur les définitions de l'impact, mériterait d'être porté au niveau du CAD. Cette étude va dans le même sens qu'un large consensus (White 2010 et 2013, World Bank, 3ie) de contestation de la définition actuelle proposée par le CAD qui définit l'impact comme un effet à long-terme. La définition de l'évaluation d'impact mériterait également d'être harmonisée en distinguant explicitement 'évaluation d'impact' et 'évaluation d'« outcomes »' bien que ces deux éléments peuvent être évalués au sein d'une même étude, comme l'exercice le démontre.

5.1.2 Faire connaître l'évaluation d'impact

Le second élément est de **faire connaître les évaluations d'outcomes et d'impact**. Les praticiens du développement paraissent être relativement familiers avec l'évaluation d'impact, mais leur connaissance reste relativement théorique ou abstraite. Car sur le terrain, le dispositif de suivi-évaluation ne permet pas l'appréciation de l'atteinte des *outcomes* ni de l'impact.

Ces évaluations ont montré que l'atteinte des *outcomes* et de l'impact est encore trop souvent considérée comme un objectif secondaire par rapport à la réalisation des « outputs ». En d'autres termes, la gestion axée sur les « outputs » prime sur celle axée sur l'atteinte des effets au niveau des bénéficiaires et à un niveau global (*outcomes* et impact).

Dans ce contexte, l'évaluation des outcomes et de l'impact est souvent négligée car mal connue tant sur le processus d'évaluation quantitative à mettre en œuvre que sur l'intérêt que cela peut apporter aux différentes parties prenantes d'une intervention. La capacité des évaluations d'impact à produire des informations très riches et utiles mérite d'être mise en avant pour convaincre les parties prenantes de les mettre en pratique.

Une évaluation quantitative des *outcomes* fournit une mesure objective et rigoureuse des effets sur les bénéficiaires en démontrant empiriquement leur causalité avec l'intervention. Une analyse qualitative rigoureuse permet de comprendre les facteurs et mécanismes ayant inhibé ou favorisé l'atteinte des effets. Elle apporte également un jugement argumenté sur l'atteinte de l'impact. De tels résultats permettent de mettre en œuvre des politiques basées sur les preuves de leur efficacité. **Les constats d'évaluation peuvent être des arguments solides pour un ministre de la coopération afin de renforcer sa position sur l'un ou l'autre dossier.** Promouvoir et faire connaître l'évaluation d'impact, c'est encourager des projets solides et convaincants, mais c'est aussi être prêt à remettre certains projets en cause sur base de preuve tangible de non effet sur les bénéficiaires.

Une bonne manière de sensibiliser et d'expliquer l'évaluation d'impact est de profiter systématiquement des présentations des rapports d'évaluation pour susciter des discussions ouvertes avec les différentes parties intéressées.

5.1.3 Prévoir un programme annuel d'évaluations d'impact en s'assurant de leur faisabilité et de leur valeur ajoutée

Chaque année, un certain nombre d'évaluations d'impact devrait être prévu et les budgets adéquats dégagés.

Le choix précis des interventions à étudier devrait faire l'objet d'une analyse explicite au cas par cas. En effet, une évaluation d'impact de qualité présente de nombreux intérêts mais représentent des budgets importants. La question de "value for money" doit être posée : d'une part, les avantages à tirer de l'évaluation d'impact en valent-ils la dépense ; d'autre part, les conditions minimales pour la réaliser sont-elles réunies ?

Les avantages d'une évaluation rigoureuse des *outcomes* et de l'impact par l'application de méthodes mixtes sont les suivants :

- Mesure des effets sur les bénéficiaires et preuve empirique de leur attribution au projet (« *outcomes* »).
- Jugement argumenté sur l'atteinte des *outcomes* non mesurables quantitativement et sur l'atteinte de l'impact (objectif global).
- Compréhension des mécanismes et éléments inhibant ou favorisant les effets d'un projet.
- Rendre compte auprès de différentes parties (bailleur, partenaire, exécutant, bénéficiaires) des effets d'une intervention. Dans le cas d'une intervention sensible ou contestée ou dans le cas d'une intervention stratégique ou simplement pour rendre compte de l'efficacité de l'aide, les constats d'une évaluation peuvent être utilisés par divers intervenants à différents niveaux de détails.
- Aide à la prise de décision. Dans le cas d'un projet pilote que l'on souhaite étendre ou généraliser ou dans le cas de projets à reconduire, l'évaluation peut guider les décisions en se basant sur des évidences démontrées scientifiquement et/ou rigoureusement argumentées.
- Capitalisation d'expériences tant sur l'atteinte des objectifs de certains projets (sommatif) qu'en termes de méthodologie de mise en œuvre de projet ou d'évaluation (formatif). Dans une logique d'apprentissage, elle se justifie lorsque l'intervention présente un caractère particulièrement innovant et un potentiel de répliquabilité élevé.

Les coûts d'une évaluation d' *outcomes* et d'impact par la mise en œuvre d'une méthodologie mixte sont relativement élevés. En effet, en plus d'un coût de collecte des données parfois important, elles prennent plus de temps (temps pour collecte des données de première main et temps pour les analyses quantitatives). Les coûts d'une évaluation d'impact sont liés à la méthodologie à mettre en œuvre et donc aux types et à la quantité de données à utiliser. Par conséquent, si des mesures d'*outcomes* pertinentes n'ont pas été collectées et stockées sur un support exploitable via le dispositif de suivi-évaluation, les coûts d'une étude d'impact seront plus élevés. **Le coût d'une évaluation ex post dépend donc en partie de la qualité de la formulation du dispositif de suivi-évaluation et de sa mise en œuvre effective.**

De même, le **contexte dans lequel se déroule l'évaluation a un effet sur les coûts de collecte d'informations** : les déplacements géographiques, la sécurité, la disponibilité des personnes à interroger, la disponibilité de personnes locales relais pour former des équipes d'enquêteurs et assurer un minimum de logistique, la difficulté de collecter certains types de données, etc. sont autant d'éléments ayant une influence significative sur le coût des données à collecter (coût résultant en partie de la durée des missions de terrain).

Par ailleurs, ces éléments influençant le coût, comme le temps à passer sur le terrain, peuvent également amener à revoir la méthodologie proposée : se concentrer sur une région (ce fut le cas au **Sénégal** et en **RDC**), annuler la collecte de données de première main (comme pour les projets en **RDC** et au **Rwanda**), ne pas collecter certains

indicateurs (abandon de la collecte de données de productivité des terres au **Maroc**), restreindre l'appréciation quantitative sur un aspect du projet (comme cela a été le cas pour **le Sénégal** –accès à l'eau via les nouveaux forage-).

La question de la « **value for money** » touche plus fondamentalement à la place de l'évaluation et donc de l'évaluation d'impact dans le système actuel de la coopération belge. Tant que l'intérêt et donc le besoin ne sont pas reconnus, il est peu probable que l'analyse coûts/bénéfices soit positive.

En plus d'une analyse coûts/bénéfices pour cerner la valeur ajoutée d'une évaluation, **il convient de s'assurer qu'elle soit faisable**. Il faut donc tenir compte de deux autres éléments :

- **La nature de l'intervention et le contexte de sa mise en œuvre sur le terrain.** Une évaluation d'impact n'aura pas de raison d'être si le cadre logique du projet présente des objectifs confus, peu cohérents ou irréalistes avec les actions et moyens mis en œuvre sur le terrain, ou si l'implémentation du projet a posé de nombreux problèmes, ou encore si l'environnement socio-politique au moment du projet était défavorable. Dans de telles circonstances, l'intervention projetée n'a pratiquement aucune chance d'avoir des effets. Il est donc inutile de dépenser des sommes pour mettre en place un dispositif d'évaluation quantitative (ceci s'est révélé être le cas pour l'évaluation du projet AETP2 en **RDC**).
- **L'environnement socio-politique au moment de l'évaluation.** Le contexte dans lequel doit se dérouler l'évaluation d'impact est également une dimension importante car il peut soit rendre impossible la collecte des informations nécessaires, soit la rendre très coûteuse, ou les deux. L'environnement politique peut ruiner une étude d'impact. En effet, le partenaire peut être réticent à voir l'efficacité d'une intervention évaluée parce qu'une évaluation d'impact mitigée constituerait un jugement négatif relatif à ses propres politiques, ou parce que l'intervention comporte des enjeux qui la dépassent et qui ont une importance cruciale pour le partenaire. **La dimension et les enjeux politiques d'une évaluation rigoureuse des effets d'un projet sont à considérer** (cette dimension a vraisemblablement été sous-estimée dans le cas au **Rwanda** pour l'évaluation du projet LRSP).

5.1.4 Garantir la qualité des évaluations

L'évaluation d'impact est un exercice complexe qui combine une perspective de recherche et une perspective de développement. C'est ce mélange qui constitue la richesse, la valeur et finalement l'utilité. Il est important de maintenir cette complémentarité et d'en tenir compte dans la rédaction des termes de référence et dans l'encadrement de l'évaluation.

Les termes de référence doivent impérativement **fixer les objectifs poursuivis** par l'évaluation ; **prévoir une mission exploratoire** ; et **envisager l'adaptation de l'évaluation en fonction des réalités du terrain**. La qualité d'une évaluation d'impact dépend de facteurs internes à l'équipe en charge mais aussi de facteurs externes imprévisibles mais déterminants, comme par exemple le refus des autorités nationales de soutenir l'évaluation.

Des échanges permanents avec le commanditaire de l'évaluation et une certaine souplesse de sa part sont indispensables pour produire un travail utile et de qualité. Par exemple, en **RDC**, le changement assez fondamental de la méthode (et de l'équipe) a permis malgré les contraintes de produire une évaluation rigoureuse.

Etant donné le recours à des méthodes statistiques relativement complexes, il est important d'être en mesure de valider des choix méthodologiques et l'appréciation de la solidité des conclusions en fonction de la qualité des données, de la qualité du contrefactuel et de la qualité des traitements statistiques effectués. Cela peut être

réalisé en interne, si les compétences existent. Si ce n'est pas le cas, il faut alors confier cette validation à un expert externe.

5.2 Réaliser des évaluations d'impact rigoureuses et de qualité

Réaliser des évaluations d'*outcomes* et d'impact rigoureuses et de qualité en appliquant une méthodologie mixte sur base de la logique d'intervention.

Cette étude a montré tout l'intérêt de mettre en œuvre une méthodologie mixte combinant des méthodes qualitatives et quantitatives dans le but d'apprécier les *outcomes* et l'impact d'une intervention. L'expérience menée confirme la littérature actuelle sur le sujet : il ne s'agit pas de choisir l'une ou l'autre méthode, **il convient de combiner de façon subtile et pertinente les méthodes quantitatives et qualitatives dans un même processus d'évaluation**. Tant la maîtrise technique (dimension quantitative) que la bonne compréhension de la logique d'intervention, du contexte, et de la thématique traitée par l'intervention (dimension qualitative) sont indispensables pour réaliser des évaluations d'impact de qualité.

Grâce à la combinaison des méthodes qualitatives et quantitatives, l'évaluation est en mesure de quantifier les effets d'un projet sur les bénéficiaires (« *outcomes* ») en démontrant leur attribution. Elle permet de porter un jugement argumenté sur l'atteinte de l'impact global d'une intervention. In fine, elle apporte une compréhension des mécanismes et des éléments de contexte qui ont favorisé ou inhibé ces effets.

Rappelons que la décision de faire une évaluation d'impact doit se fonder sur une analyse préalable de sa pertinence et de sa faisabilité, tant sur le terrain qu'en termes financiers (cf. supra 5.1.3). **Il convient également de réfléchir à sa faisabilité sur le plan méthodologique.**

La méthodologie d'une évaluation d' *outcomes* et d'impact est influencée par les objectifs poursuivis de l'évaluation (rendre compte, aider à la décision et/ou capitaliser les expériences), par le moment de sa mise en œuvre (en même temps que la mise en œuvre du projet, après le démarrage de celui-ci ou encore lorsqu'il est terminé) et par les moyens financiers disponibles.

5.2.1 Concevoir l'évaluation au moment de la conception du projet

Il est recommandé de concevoir la méthodologie d'évaluation des *outcomes* et de l'impact d'un projet en même temps que sa conception (ex ante). Cela présente un double avantage :

- Tout d'abord, la conception d'un processus d'évaluation d'impact au même moment que la conception du projet peut avoir un effet sur la qualité de la conception du projet lui-même. Une conception de l'analyse de l'impact avant la mise en œuvre d'un projet oblige les praticiens de l'évaluation et du développement à formuler concrètement les hypothèses sous-jacentes à la réalisation des « *outputs* », à l'atteinte des *outcomes* et de l'impact. Cela nécessite une meilleure connaissance du contexte ainsi qu'une analyse de risques approfondie. Cette analyse rigoureuse ex ante pourrait mener à revoir certaines activités ou la manière de les mettre en œuvre afin de maximiser les chances de réussite de l'intervention. Une meilleure compréhension des enjeux, des risques et du contexte peut également mener à réviser les objectifs de l'intervention pour les rendre plus réalistes (modifier les valeurs cibles des indicateurs par exemple).

En bref, une conception du dispositif évaluatif au moment de la conception du projet permet d'utiliser le cadre logique comme un véritable outil de gestion et d'évaluation.

- Ensuite, concevoir un système d'évaluation relativement tôt dans le processus d'une intervention permet la mise en œuvre d'une méthodologie plus rigoureuse qui accroît de façon significative la qualité de l'évaluation quantitative, ce qui implique de :
 - (i) Réaliser une « *baseline* » pour des indicateurs pertinents et bien définis auprès d'un échantillon de bénéficiaires et auprès d'un contrefactuel ;
 - (ii) Mettre en place un contrefactuel, en encourageant une mise en œuvre séquentielle du projet (« *pipeline process* ») ou sur base de critères objectifs réfléchis au moment de l'intervention afin qu'il soit crédible ;
 - (iii) Collecter des données quantitatives de suivi sur les indicateurs tout au long de l'intervention et en fin d'intervention, auprès d'un groupe de bénéficiaires et idéalement auprès du contrefactuel et définir les modalités de collecte et la fréquence avec précision.

Compte tenu de la place actuelle de l'évaluation au sein de la coopération bilatérale belge, il est peu probable que le processus d'évaluation des *outcomes* et de l'impact soit élaboré et effectif sur le terrain dès la mise en œuvre d'un projet. Cependant, les évaluations réalisées dans le cadre de cette étude montrent qu'il est toutefois possible d'utiliser de méthodes quantitatives même lorsque les conditions optimales ne sont pas toutes réunies. **Il est donc recommandé de faire usage des méthodes quantitatives d'analyse tant que cela est faisable et pertinent que l'on soit dans une conception ex ante ou ex post du processus d'évaluation des *outcomes* et de l'impact.**

Enfin, signalons que ne pas procéder à des évaluations d'impact sous prétexte qu'on reproduit une intervention dont les preuves de succès ont été démontrées n'est pas satisfaisant. En effet, les évaluations d'impact ont une validité externe limitée, le mécanisme d'aide et le contexte sont déterminants dans la réussite d'un projet, deux éléments difficilement imitables (cf. supra pour plus d'explications). **Il est donc recommandé de procéder à l'évaluation des projets par une méthodologie mixte moyennant les respects des conditions nécessaires et minimales.**

5.2.2 Etapes successives et éléments importants pour mener une évaluation d'impact ex post rigoureuse

La mise en œuvre d'un dispositif d'évaluation combinant des méthodes mixtes implique la mobilisation d'une équipe d'experts rodés en évaluation qualitative et dans l'application de méthodes quantitatives.

Cependant, d'autres éléments sont également importants :

- Avoir un partenaire local qualifié, connaissant à la fois les réalités du terrain et comprenant les exigences méthodologiques.
- Trouver un accompagnateur sur le terrain connaissant bien le projet et la zone en évitant les conflits d'intérêt. Son rôle doit être bien défini au départ, afin qu'il n'influence pas les répondants.
- Prévoir l'implication des différentes parties prenantes au projet dans le processus d'évaluation. Il est particulièrement important de les impliquer dans la phase préparatoire de l'évaluation, ainsi que lors de la restitution des constats de l'évaluation. Cela permet d'adresser les questions les plus pertinentes pour les

5. Recommandations

acteurs en jeu, mais aussi de leur faire valider les constats. Ce type de démarche permet d'accroître la qualité et la pertinence de l'étude. Par ailleurs, c'est aussi l'occasion d'échanger sur le processus méthodologique et ainsi de les sensibiliser à l'évaluation d'*outcomes* et d'impact.

Ci-dessous, sont exposées **les étapes du dispositif évaluatif à mettre en œuvre pour réaliser des évaluations ex post rigoureuses et de qualité.**

- (1) **Phase préparatoire.** Les objectifs sont de reconstruire le cadre logique, d'identifier la méthodologie quantitative à appliquer, et de préparer les outils de collecte de données.
 - **Étude documentaire pour reconstruire la logique d'intervention :** analyse des documents existants pertinents, analyse du contexte socio-économique, politique et culturel au moment du projet et au moment où doit se dérouler l'évaluation ; analyse de la ou des thématique(s) traitée(s) dans le projet (via, entre autres, la littérature académique et scientifique) et rencontres avec certaines personnes clés de l'intervention présentes en Belgique.
 - **Formulation des hypothèses sous-jacentes aux liens de causalité menant aux *outcomes* et à l'impact dans la logique d'intervention et définition des indicateurs pertinents.** L'idée est ici de mettre en évidence les relations causales entre les « inputs », les « outputs », les *outcomes* et l'impact. Cet exercice permet de se rendre compte d'éventuels problèmes dans la chaîne des effets attendus. Par exemple, l'identification d'*outcomes* et/ou d'impact non réalistes avec les actions et moyens mis en œuvre, l'identification d'éléments du contexte propices ou, au contraire, néfastes au bon déroulement des effets. Il convient alors de reformuler certains *outcomes* et/ou un impact plus cohérents avec la logique d'action. C'est à ce stade qu'il faut critiquer la pertinence et la qualité des indicateurs disponibles dans le système de suivi-évaluation du projet. Finalement, il convient d'établir une liste d'indicateurs pertinents à collecter sur le terrain afin d'être en mesure d'apprécier les *outcomes* et l'impact du projet.
 - **Identification des données de seconde main et élaboration des outils de collecte d'informations.** Sur base des hypothèses à tester et des indicateurs à collecter, il convient de chercher les données pertinentes qui seraient déjà disponibles dans les archives du projet ou via des statistiques nationales. En fonction de ce qui doit être complété, il convient ensuite d'élaborer des outils de collecte de données : des guides d'entretien (pour des enquêtes de groupe ou des entretiens individuels avec certaines personnes clés de l'intervention), et/ou un questionnaire pour une enquête de ménages. Dans une situation ex post, il faut tenter de reconstruire la situation pré-intervention pour les bénéficiaires et le contrefactuel via des données faisant appel à la mémoire des répondants ou via des données de seconde main.
 - **Identification d'un contrefactuel** pour être en mesure de démontrer l'attribution des effets au projet. Sur base d'une bonne connaissance du contexte et de la zone d'intervention, il convient de sélectionner un groupe de non-bénéficiaires qui présente des caractéristiques similaires au groupe des bénéficiaires avant l'intervention. Cette étape est difficile mais pas impossible. Il faut cependant être conscient de deux problèmes majeurs : l'effet de contamination ou contagion (« *spill over effect* ») et le problème de biais de sélection (cf. supra dans la méthodologie pour une explication détaillée).
 - **Définition des échantillons et de la zone d'enquête.** Sur base des objectifs poursuivis par l'étude d'impact (rendre compte, aide à la décision, capitalisation), des contraintes temporelles et financières et sur base des données de seconde main déjà disponibles, il convient de définir une taille d'échantillon suffisante pour s'assurer que l'on détecte, avec un certain degré de précision, les « *outcomes* », le cas échéant. La sélection de la zone d'enquête, ainsi que celle des répondants doit être au maximum aléatoire ou se faire sur base de critères objectifs sans

influence du partenaire ou des exécutants du projet (sans quoi cela pourrait biaiser les résultats).

- **Mission exploratoire.** C'est d'abord l'occasion de s'assurer d'avoir l'aval des autorités locales sur la conduite d'une évaluation et leur support éventuel pour la réaliser. Ensuite, c'est l'opportunité de rencontrer certaines personnes clés de l'intervention vivant dans le pays et de se rendre compte des « outputs » réalisés sur le terrain. Des échanges sur les questions pertinentes à traiter peuvent également avoir lieu. C'est également l'occasion de collecter certaines données de seconde main et de s'assurer de leur fiabilité. C'est aussi le moment où l'on teste les différents outils de collecte d'informations, et qu'on prépare la logistique de la phase de collecte de données.

Les informations collectées pendant la mission exploratoire permettent de réviser la logique d'intervention à la lumière des réalités du terrain, de revoir les hypothèses à tester, ainsi que d'affiner les indicateurs à collecter en apportant des corrections aux outils de collecte d'informations.

- (2) **Phase terrain.** Les objectifs sont de collecter des données de bonne qualité en quantité suffisante ; de rassembler un maximum d'observations et d'informations également via des échanges informels. Ce moment relativement long sur le terrain est aussi souvent l'occasion de répondre aux questions des bénéficiaires et autres parties prenantes au projet, de soulever des problèmes et de débattre de façon informelle. Il convient donc d'avoir une équipe d'enquêteurs bien préparée et être attentif à ce que chacun reste dans son rôle.

- **La collecte de données de première main** à proprement parlé ne commence qu'après une phase de test des questionnaires. Cette étape est cruciale pour garantir la qualité des données collectées sur le terrain. Elle consiste à s'assurer que les enquêteurs maîtrisent l'outil et qu'ils sont compétents pour la tâche demandée. Cela permet aussi de revoir la formulation de certaines questions et d'affiner encore la connaissance des réalités du terrain.

- La mise en œuvre des différents outils de collecte de données se réalise **sous la supervision** d'un évaluateur qui accompagne l'équipe d'enquêteurs sur le terrain. La présence de l'évaluateur durant la phase d'enquête présente plusieurs avantages : faire face aux imprévus du terrain (par exemple le climat ou l'état des routes qui empêche de se rendre dans certaines zones), s'assurer que les enquêteurs interrogent bien les gens sélectionnés, corriger des questionnaires au fur et à mesure, y apporter des modifications le cas échéant. De plus cela permet à l'évaluateur, par l'observation et les rencontres informelles, de se forger une meilleure compréhension du contexte et des enjeux locaux.

- Finalement, il faut **encoder les données** dans un support informatique exploitable (par exemple CSprou). Pour des raisons de coût de main d'œuvre, il est conseillé d'encoder les questionnaires localement mais il faut s'assurer d'avoir un partenaire compétent pouvant assurer le suivi et la qualité de l'encodage (par exemple un chercheur d'université). Notons que la technique du double encodage qui consiste à encoder un même questionnaire par deux personnes différentes, permet de rapidement identifier des erreurs d'encodage. Cependant, cette méthode est onéreuse. Si le support d'enquête est une tablette, la phase d'encodage n'existe plus. Les données sont automatiquement disponibles. Néanmoins, l'utilisation de tablette présente certains inconvénients (recharge dans des zones sans électricité, utilisation par des personnes peu formées à l'informatique, retour en arrière lors d'incohérences, etc.).

- (3) **Phase d'analyse.** Les objectifs sont de faire la synthèse des résultats obtenus par l'analyse quantitative et qualitative, d'en démontrer les liens avec le projet en expliquant les éléments du contexte et les mécanismes ayant favorisé (ou inhibé en cas de constat de non effet) ces effets.

5. Recommandations

- **Nettoyage des données.** Selon la qualité de l'encodage, cette phase est plus ou moins conséquente. Seule une personne connaissant bien le terrain et la phase d'enquête est en mesure de faire ce travail. En effet, il s'agit de s'assurer de la cohérence des données encodées, cohérence entre les différentes parties d'un même questionnaire mais aussi en cohérence avec les observations faites sur le terrain.
- **Analyse des données via l'utilisation de méthodes quantitatives.** Dans un premier temps, il convient de tester les hypothèses sous-jacentes aux liens de causalité menant aux « outcomes », en quantifiant les effets observés et en démontrant leur attribution au projet. Pour se faire, on utilise les données quantitatives (première et seconde main) pour réaliser des tests statistiques et économétriques (sur le logiciel STATA par exemple) mais on utilise également les données qualitatives pour argumenter certains liens de causalité et apprécier certains effets non quantifiables. Dans un deuxième temps, il s'agit d'apporter des explications sur les résultats obtenus, c'est-à-dire expliquer les mécanismes ayant mené aux effets ou à l'absence d'effet du projet. Une analyse quantitative offre un potentiel large d'apprentissages sur ce qui fonctionne et ne fonctionne pas dans l'aide au développement.

- (4) **Phase de restitution.**²⁵ L'objectif poursuivi est double. D'une part, faire connaître l'évaluation d'impact en partant d'un cas concret tant en termes de constats sur les effets des projets qu'en termes de méthodologie d'évaluation. D'autre part, faire valider les constats et/ou en débattre avec les personnes concernées.

Une fois les rapports d'évaluation rédigés et approuvés par le commanditaire, il est intéressant d'organiser une restitution sur le terrain auprès du partenaire, des exécutants et des bénéficiaires du projet. C'est une façon de se rendre compte de la valeur ajoutée d'une évaluation d'impact de qualité par rapport à un autre type d'évaluation et de soulever de véritables débats sur l'efficacité de l'aide.

5.3 « Management towards outcomes and impact » pour des projets plus réalistes visant l'atteinte d'effets tangibles

Intégrer l'évaluation des *outcomes* et de l'impact dans la gestion des projets afin de concevoir des projets plus réalistes qui visent en priorité l'atteinte des *outcomes* et de l'impact. Cela consiste à intégrer l'évaluation dès la formulation d'un projet et dans sa mise en œuvre, ainsi qu'à élaborer un dispositif de suivi-évaluation efficace pour mesurer les *outcomes* et apprécier l'impact d'un projet.

La présente analyse a montré les faiblesses des trois projets au niveau de leur conception, de leur mise en œuvre et de leur suivi-évaluation. Ces faiblesses expliquent au moins en partie le peu d'impacts positifs constatés dans les trois projets examinés.

Les projets ne se fixent pas comme priorité l'atteinte des « outcomes », ou de l'impact mais se focalisent plutôt sur les réalisations physiques et les décaissements financiers. Dans un contexte de recherche d'amélioration de l'efficacité de la dépense publique et d'amélioration de l'efficacité de l'aide au développement, ce constat est interpellant.

A l'issue de cet exercice d'évaluation d'impact ex post de quatre projets de développement et sur base des conclusions obtenues, le rapport propose une piste pour améliorer la conception/formulation, la mise en œuvre et le suivi-évaluation des

²⁵ Dans cet exercice, une restitution en Belgique est planifiée, cependant une restitution sur le terrain n'est pas prévue.

interventions financées par la coopération belge, dans le but d'améliorer l'atteinte des *outcomes* et de l'impact.

Il est recommandé tout d'abord **d'intégrer l'évaluation des *outcomes* et de l'impact dès la conception des projets, ainsi que dans leur mise en œuvre et leur suivi.** Cela ne signifie pas qu'il faille réaliser des évaluations d'impact pour chaque projet. Cela implique de se focaliser sur l'atteinte des objectifs spécifiques et de l'objectif global dès le départ, et non de se préoccuper uniquement des réalisations physiques et du décaissement du budget en temps et en heure.

L'intégration de l'évaluation des *outcomes* et de l'impact dans la gestion des projets présente deux avantages.

- Il nécessite de **concevoir le processus d'évaluation au moment de la conception/formulation d'un projet.** Cela augmente de façon significative la qualité de l'évaluation. Plus particulièrement cela permet de concevoir un dispositif d'évaluation rendant possible l'application de méthodes quantitatives. (cf. supra recommandation 5.2.1)
- Il peut influencer la manière de concevoir et de mettre en œuvre des projets, et ainsi **tendre vers une plus grande efficacité de l'aide.** En effet, l'intégration de l'impact dès la conception d'un projet aura pour effet de ne pas entreprendre et/ou de suspendre des projets irréalistes. Cela aura également pour conséquence d'utiliser le cadre logique comme véritable outil de gestion, de suivi et d'évaluation interne.

Afin de rendre cela possible, il est recommandé de créer de bons incitants à tous les niveaux (DGD, CTB, partenaires, bénéficiaires).

5.3.1 Intégrer l'évaluation d'impact dès la conception et renforcer le processus de validation des projets pour ne pas démarrer des projets irréalistes

Intégrer l'évaluation d'impact dès la conception permet de réaliser un travail de réflexion approfondi sur les facteurs pouvant inhiber ou favoriser le succès d'une intervention. Cela nécessite de formuler les hypothèses sous-jacentes à l'ensemble de la chaîne de causalité de la logique d'action d'un projet, et par conséquent permet une meilleure connaissance du contexte spécifique, des enjeux à tous les niveaux et des risques. Un tel travail préalable permet alors de concevoir/formuler des projets en cohérence avec les réalités du terrain.

Les évaluations réalisées montrent que plusieurs éléments touchant à la conception d'un projet sont à améliorer :

- La connaissance du contexte, des enjeux à différents niveaux (politique, social, économique, en amont et en aval du secteur touché par le projet) et leur incidence sur le bon déroulement du projet ;
- La formulation d'objectifs globaux et spécifiques réalistes, bien définis, quantifiés (valeur cible à atteindre) et mesurables par des indicateurs dont on connaît la valeur initiale ;
- L'analyse des risques et une appréciation objective de leur incidence potentielle sur le projet ;
- La revue systématique de la littérature (y compris académique) pour s'assurer que les leçons apprises ailleurs soient prises en compte.

Intégrer l'évaluation d'impact dès la conception est une façon d'améliorer la conception d'un projet, mais parallèlement à cela, il est nécessaire de **renforcer le processus de validation des projets avant leur lancement.** Même si ce processus existe sous la forme d'un comité de contrôle de qualité, il n'est pas suffisamment efficace, car il laisse passer des projets dont la probabilité qu'ils atteignent les objectifs escomptés est faible. En effet, une étude approfondies ex ante l'aurait montré pour les quatre projets évalués.

5. Recommandations

Cela demande une volonté forte, des moyens suffisants pour la préparation des actions, une organisation adéquate entre la DCG et la CTB, mais surtout la mise en place de bons incitants afin que chaque partie œuvre à cette amélioration.

5.3.2 Mettre en pratique une véritable gestion axée sur les résultats (« outputs mais aussi *outcomes* et impact »)

Dans les projets examinés, les dispositifs de suivi-évaluation ont été efficaces pour le suivi opérationnel de la réalisation des outputs, mais aucune donnée pertinente exploitable pour apprécier les *outcomes* et l'impact des projets n'a été générée par ce dispositif. Même face à des difficultés au niveau de l'exécution des activités, **les décisions ont toujours privilégié le décaissement**. Et cela, même lorsqu'il paraissait évident que la réalisation de ces « outputs » ne pouvait conduire aux *outcomes* attendus.

Intégrer l'évaluation d'*outcomes* et de l'impact au moment de la conception du projet et de sa mise en œuvre permet de mettre en place un véritable dispositif de gestion axée sur les résultats. Il convient donc de concevoir un système de suivi-évaluation efficace et opérationnel pour faire le suivi de la réalisation des « outputs », mais également le suivi et l'évaluation de l'atteinte des *outcomes* (et de l'impact). Il ne suffit pas d'acter des réalisations, il faut également rester critique quant aux effets des « outputs » réalisés (sont-ils utilisés ? par qui ? pour quoi faire ?).

Mettre en place une **gestion axée sur les *outcomes* et l'impact (« management towards results »)** nécessite un dispositif de suivi-évaluation permettant

- De rendre compte régulièrement de l'avancée du projet (état des réalisations et décaissements) et des causes des retards ou problèmes éventuels ;
- De rendre compte des manquements au niveau de l'atteinte des *outcomes* et d'en identifier les causes ;
- De faire remonter l'information concernant les faits préoccupants sur le projet ou des inquiétudes face à l'atteinte des objectifs fixés ;
- De disposer d'une capacité de décision pour remettre en cause ou revoir un projet qui n'est plus en ligne avec les *outcomes* escomptés et l'impact attendu ; et donc potentiellement décider de suspendre ou annuler une intervention le cas échéant.

Pour qu'un **système de suivi-évaluation efficace soit opérationnel**, il est nécessaire de définir avec précision les **modalités de collecte d'information** : qui collecte quelles données, à quelle fréquence, sur quel support, avec quel soutien, dans l'objectif d'en faire quoi ? De plus, il est indispensable de **créer les bons incitants à chaque niveau** (DGD, CTB, partenaire, bénéficiaires), sans quoi, il ne sera pas opérationnel, par manque de moyen et/ou de temps et/ou de ressources humaines.

6. Bibliographie

Bamberger, M., 2012. Introduction to mixed methods in impact evaluation. Impact evaluation notes, Rockefeller Foundation & Interaction.

CAD, 2010. Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats.

Deaton A.S., 2009. Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development. Working paper 14690. National Bureau of Economics Research.

Gertler P.J. et al., 2011. L'évaluation d'impact en pratique. World Bank.

IEG (Impact Evaluation Group), 2013. Impact evaluation-the experience of the independent evaluation group of the World Bank.

IEG (Impact Evaluation Group), 2013. Conducting quality impact evaluations under budget, time and data constraints. World Bank.

Naudet J.D., Delarue J., Bernard T., 2012. Impact evaluation: a tool for accountability? Lessons from experience at AFD. AFD working paper.

White, H., 2009. Theory-based impact evaluation: principles and practice. Working paper 3, 3ie.

White, H., 2013. An introduction to the use of randomized control trial to evaluate development interventions. Working paper 9, 3ie.

Annexes

ANNEXE 1 : RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PROJET AETP2 EN RDC

ANNEXE 2 : RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PARPEBA AU SÉNÉGAL

ANNEXE 3 : RAPPORT D'ÉVALUATION EX POST DE L'IMPACT DU PROJET PMH AU MAROC

Les annexes ne sont pas imprimées et sont reprises uniquement sur le CD-ROM joint à la synthèse de l'évaluation ou sur le site internet.



ROYAUME DE BELGIQUE

Service public fédéral
**Affaires étrangères,
Commerce extérieur et
Coopération au Développement**

Service de l'Évaluation spéciale de la Coopération internationale

Egmont • rue des Petits Carmes 15, B-1000 Bruxelles • + 32 2 501 38 34 • www.diplomatie.belgium.be • www.dg-d.be • ses-dbe@diplobel.fed.be

Dépôt légal n° 0218/2014/002