



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

Service de l'Évaluation spéciale de la Coopération internationale

Méta-évaluation des programmes des acteurs non gouvernementaux



Rapport final

Juillet 2013

La présente évaluation a été réalisée par ADE 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

Juillet 2013

Le présent rapport doit être cité comme suit: Service de l'Évaluation spéciale (2013), *Méta-évaluation des programmes des acteurs non gouvernementaux*, SPF Affaires étrangères, Commerce extérieur et Coopération au développement, Bruxelles

Production graphique : Sandra Fernandez Bernardo, Service Communication SPF

Impression : Service Imprimerie

Evaluation N° S4/2012/01

Dépôt légal : 0218/2013/014

Ce document est disponible en format PDF en français et en néerlandais sur le CD-ROM joint en annexe du résumé 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 méta-évaluation. Nous remercions en particulier les nombreux membres du Comité d'accompagnement et les personnes rencontrées pour le temps et l'énergie consacrés à cette étude d'un type particulier, ainsi que les personnes enquêtées pour l'information fournie.

Corrigendum (juillet 2013): ATOL a été accidentellement incluse dans la catégorie des 'autres partenaires' au lieu de celle des ONG. L'effet sur les analyses chiffrées par catégorie d'acteurs est néanmoins limité et ne change en rien les constats de cette étude.

Table des matières

TABLE DES MATIÈRES	5
LISTE DES ENCADRÉS, DES FIGURES ET DES TABLEAUX	7
ACRONYMES ET ABRÉVIATIONS.....	9
GLOSSAIRE.....	11
RÉSUMÉ.....	13
1. INTRODUCTION	17
1.1 OBJET, OBJECTIFS ET CHAMP	17
1.2 STRUCTURE DU RAPPORT	18
2. MÉTHODOLOGIE.....	19
2.1 PROCESSUS D'ÉVALUATION	19
2.2 ÉCHANTILLON	20
2.3 QUESTIONS D'ÉVALUATION.....	20
2.4 OUTILS D'ÉVALUATION.....	21
2.5 DÉFIS ET LIMITES	23
3. RÉPONSES AUX QUESTIONS D'ÉVALUATION	27
3.1 POLITIQUES ET PRATIQUES DE LA DGD ET DES ANG – Q1	28
3.1.1 Politique et pratiques de la DGD (Q1.1 et Q1.2).....	29
3.1.2 Approche des ANG (Q1.2 et Q1.3).....	35
3.2 QUALITÉ DES ÉVALUATIONS – Q2.....	38
3.2.1 Qualité de la demande et couverture de cette demande (Q2.1).....	39
3.2.2 Qualité des rapports d'évaluation (Q2.2).....	41
3.2.3 Facteurs explicatifs de la qualité (Q2.1, Q2.2 et Q2.3)	48
3.3 INDEPENDANCE – Q3.....	52
3.3.1 Indépendance dans le lancement des évaluations (Q3.1)	53
3.3.2 Indépendance dans la conduite des évaluations (Q3.2).....	55
3.3.3 Indépendance des rapports d'évaluation (Q3.3)	55
3.4 CREDIBILITE – Q4	57
3.5 CONTENU DES EVALUATIONS – Q5	62

3.5.1	Remarques préliminaires.....	63
3.5.2	Conclusions récurrentes.....	64
3.5.3	Recommandations récurrentes	66
3.6	UTILISATION PAR LES ANG – Q6.....	68
3.6.1	Capitalisation sur les connaissances et partage de leçons (Q6.1).....	69
3.6.2	Reddition de compte (Q6.1)	71
3.6.3	Prise de décision (Q6.3).....	72
3.7	UTILISATION PAR LA DGD – Q7	74
3.7.1	Directives sur l'utilisation et Réception des rapports (Q7.1 et Q7.2)	75
3.7.2	Utilisation des évaluations par la DGD (Q7.1 et Q7.2).....	76
3.7.3	Organisation interne (Q7.3)	77
4.	CONCLUSIONS	81
4.1	APPRÉCIATION GÉNÉRALE	81
4.2	CONCLUSIONS	82
5.	RECOMMANDATIONS	91
	ANNEXES	107
	RÉPONSE MANAGÉRIALE DE LA DGD.....	109
	RÉPONSE MANAGÉRIALE DES ANG	115

Les annexes de cette méta-évaluation sont reprises dans un volume séparé, disponible sur CD-ROM, auprès du Service de l'Évaluation spéciale ou sur le site http://diplomatie.belgium.be/fr/politique/cooperation_au_developpement/evaluation

Liste des encadrés, des figures et des tableaux

LISTE DES ENCADRÉS

ENCADRÉ 1 – COÛTS DE GESTION ET LIMITE DU NIVEAU D'HONORAIRES	33
ENCADRÉ 2 – DISTINCTION PEU FAITE ENTRE TROIS TYPES D'OBJECTIFS	34
ENCADRÉ 3 – FORMATIONS SUIVIES EN ÉVALUATION.....	36
ENCADRÉ 4 – POLITIQUES D'ÉVALUATION D'ANG INTÉRESSANTES.....	37
ENCADRÉ 5 – SURESTIMATION DE LA QUALITÉ DES ÉVALUATIONS PAR LES ANG.....	48
ENCADRÉ 6 – ACCOMPAGNEMENT STRUCTUREL	51
ENCADRÉ 7 – IDENTIFICATION DES ÉVALUATEURS	59
ENCADRÉ 8 – MOMENT DE L'ÉVALUATION	73
ENCADRÉ 9 – RELATION ENTRE LA DGD ET LES ANG	77

LISTE DES FIGURES

FIGURE 1 – OBJET, OBJECTIFS ET CHAMP	17
FIGURE 2 – PROCESSUS D'ÉVALUATION	19
FIGURE 3 – OUTILS D'ÉVALUATION	21
FIGURE 4 – PRATIQUES D'ÉVALUATION INNOVANTES OBSERVÉES.....	45
FIGURE 5 – APPRÉCIATION DES RÉPONDANTS DE LA DGD SUR DIFFÉRENTS CRITÈRES DE QUALITÉ	61
FIGURE 6 – THÉMATIQUES TRAITÉES PAR LES CONCLUSIONS ET RECOMMANDATIONS DES RAPPORTS.....	63
FIGURE 7 – UTILISATION DES RAPPORTS PAR LES ANG	69
FIGURE 8 – DIFFUSION DES RAPPORTS D'ÉVALUATION	70
FIGURE 9 – RÉCEPTION DES RAPPORTS D'ÉVALUATION.....	76
FIGURE 10 – CONCLUSIONS	82
FIGURE 11 – RECOMMANDATIONS	91
FIGURE 12 – PRIORITÉ DES RECOMMANDATIONS	92

LISTE DES TABLEAUX

TABLEAU 1 – OUTILS D'ÉVALUATION, PAR QUESTION D'ÉVALUATION	22
TABLEAU 2 – APERÇU DES QUESTIONS D'ÉVALUATION	27
TABLEAU 3 – QUALITÉ DES RAPPORTS D'ÉVALUATION, PAR TYPE D'ANG	41
TABLEAU 4 – APPRÉCIATION DE LA QUALITÉ DES 66 RAPPORTS D'ÉVALUATION (SYNTHÈSE SUR BASE DES FICHES RAPPORTS)	42
TABLEAU 5 – BUDGET DES ÉVALUATIONS	49
TABLEAU 6 – QUALITÉ DES RAPPORTS D'ÉVALUATION, PAR TYPE D'ÉVALUATEURS	50
TABLEAU 7 – PROCESSUS DE QUALITÉ STRUCTURÉ	60

Acronymes et abréviations

ANG	Acteur non gouvernemental
C	Conclusion
CA	Comité d'accompagnement (de l'évaluation)
CAD	Comité d'aide au développement (de l'OCDE)
DGD	Direction générale de la Coopération au développement (belge)
I	Indicateur
M&E	Suivi (monitoring) et évaluation
OCDE	Organisation de coopération et de développement économiques
ONG	Organisation non gouvernementale
Q	Question d'évaluation
R	Recommandation
SES	Service de l'Évaluation spéciale de la Coopération internationale
SPF	Service public fédéral
TdR	Termes de référence

Glossaire

Ce glossaire est une liste non exhaustive de termes utiles pour cette méta-évaluation. Les définitions sont celles du Comité d'aide au développement (CAD) de l'Organisation de coopération et de développement économiques (OCDE), telles que publiées dans le Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats (2002, réimprimé sans changement en 2010)¹.

Cadre logique Outil de gestion visant à améliorer la conception des actions, le plus souvent au niveau des projets. Il suppose la détermination des éléments stratégiques (ressources, extrants [*produits*], réalisations, impacts) et de leurs relations causales, des indicateurs ainsi que des facteurs extérieurs (risques) qui peuvent avoir une influence sur le succès ou l'échec de l'action. Il facilite ainsi la planification, l'exécution et l'évaluation d'une intervention de développement.

Efficacité Mesure selon laquelle les objectifs de l'action de développement ont été atteints, ou sont en train de l'être, compte tenu de leur importance relative.

Remarque: terme également utilisé comme système de mesure globale (ou comme jugement) du mérite et de la valeur d'une activité ; mesure selon laquelle une intervention a atteint, ou est en train d'atteindre, ses principaux objectifs pertinents, de façon efficiente et durable, et avec un impact positif en terme de développement institutionnel.

Efficience Mesure selon laquelle les ressources (fonds, expertise, temps, etc.) sont converties en résultats de façon économe.

Évaluation Appréciation systématique et objective d'un projet, d'un programme ou d'une politique, en cours ou terminé, de sa conception, de sa mise en œuvre et de ses résultats. Le but est de déterminer la pertinence et l'accomplissement des objectifs, l'efficience en matière de développement, l'efficacité, l'impact et la durabilité. Une évaluation devrait fournir des informations crédibles et utiles permettant d'intégrer les leçons de l'expérience dans le processus de décision des bénéficiaires et des bailleurs de fonds.

Le terme « évaluation » désigne également un processus aussi systématique et objectif que possible par lequel on détermine la valeur et la portée d'une action de développement projetée, en cours ou achevée.

Remarque : dans certains cas l'évaluation suppose la définition de normes appropriées, l'appréciation des performances par rapport à celles-ci, et en juger les résultats attendus et obtenus. Elle permet de dégager des leçons pertinentes.

Évaluation Évaluation d'une action de développement conduite par des

¹ <http://www.oecd.org/dataoecd/29/21/2754804.pdf>

indépendante	services ou des personnes non liés aux responsables de la conception et de la mise en œuvre de l'action de développement. Remarque : la crédibilité d'une évaluation dépend en partie de l'indépendance avec laquelle elle a été conduite. L'indépendance implique une liberté par rapport aux influences politiques et aux pressions des organisations. Elle est caractérisée par l'accès libre et complet à l'information et par une autonomie totale pour mener des investigations et en rendre compte.
Impact	Effets à long terme, positifs et négatifs, primaires et secondaires, induits par une action de développement, directement ou non, intentionnellement ou non.
Méta-évaluation	Évaluation conçue comme une synthèse des constatations tirées de plusieurs évaluations. Le terme est également utilisé pour désigner l'évaluation d'une évaluation en vue de juger de sa qualité et/ou d'apprécier la performance des évaluateurs.
Pertinence	Mesure selon laquelle les objectifs de l'action de développement correspondent aux attentes des bénéficiaires, aux besoins du pays, aux priorités globales, aux politiques des partenaires et des bailleurs de fonds. Remarque : rétrospectivement, la question de la pertinence consiste souvent à s'interroger sur le fait de savoir si les objectifs de l'action ou sa conception sont encore appropriés compte tenu de l'évolution du contexte.
Suivi (<i>monitoring</i>)	Processus continu de collecte systématique d'informations, selon des indicateurs choisis, pour fournir aux gestionnaires et aux parties prenantes d'une action de développement en cours, des éléments sur les progrès réalisés, les objectifs atteints et l'utilisation des fonds alloués.

Résumé

Introduction

Une méta-évaluation est une « évaluation conçue comme une synthèse des constatations tirées de plusieurs évaluations ». Le terme est également utilisé pour désigner « l'évaluation d'une évaluation en vue de juger de sa qualité et/ou d'apprécier la performance des évaluateurs ».

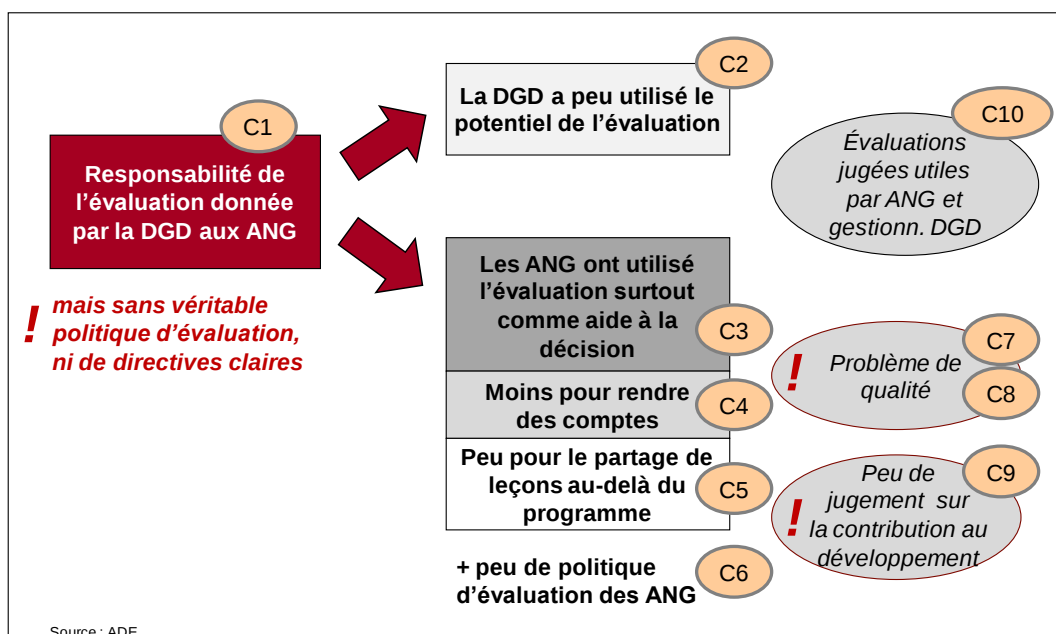
Cette méta-évaluation étudie les **politiques et pratiques d'évaluation** de la **coopération belge (DGD)** et d'**acteurs non-gouvernementaux (ANG)** dans le cadre des **programmes des ANG qui sont cofinancés par la DGD**. Elle vise en particulier à déterminer la **qualité** de l'évaluation, son **indépendance**, sa **crédibilité** et son **utilité**. Elle identifie également les **leçons récurrentes** dans l'ensemble des évaluations.

Cette étude a été commanditée par le Service de l'Évaluation spéciale, un service externe rapportant directement au Ministre de la Coopération au Développement. Elle couvre **64 ANG** soit la plupart de ceux qui avaient un programme cofinancé: **ONG, universités, institutions scientifiques, syndicats, pouvoirs locaux, etc.** Elle étudie la période de financement **2008-2011**, qui est donc *antérieure* à la réforme interne à la DGD de 2012, à la nouvelle Loi relative à la Coopération au Développement d'avril 2013 et au futur arrêté royal qui vise à redéfinir le cadre législatif et réglementaire des subventions de la DGD aux ANG pour la décennie à venir.

Cette méta-évaluation s'appuie sur une **méthodologie à la fois solide et innovante**. Celle-ci comprend une analyse systématique des rapports d'évaluation sur base de 88 indicateurs de qualité. Elle inclut également, entre autres, trois enquêtes : une auprès des ANG, une pour les équipes d'évaluation et une destinée aux gestionnaires de la DGD.

Conclusions

Les conclusions principales de l'étude sont résumées dans la figure ci-dessous. Elles sont décrites par la suite.



L'étude montre qu'il n'y avait **pas à la DGD de véritable politique d'évaluation** fixant des objectifs, définissant des modalités et prévoyant les ressources nécessaires. Celle-ci s'est largement limitée à transférer la responsabilité de l'évaluation aux ANG eux-mêmes. Ceci alors que la coopération non gouvernementale belge représentait pas moins de 901 millions € ou 18% du total de l'aide de la DGD durant les quatre années étudiées. (C1)

La **DGD** n'a dès lors **que peu pu utiliser le potentiel de l'évaluation pour ses propres besoins**, que ce soit pour rendre compte des résultats auprès des contribuables et des destinataires, pour aider à sa prise de décision, ou pour capitaliser et partager les leçons tirées des programmes. Elle ne disposait par exemple pas d'une part significative des évaluations, suite au fait que le cadre réglementaire ne prévoyait pas que les ANG dussent les lui transmettre. (C2)

Les **ANG** ont largement apprécié la flexibilité offerte. Ils ont principalement fait réaliser des évaluations comme un **outil d'aide à la décision**, pour disposer d'une base solide pour revoir l'approche du programme en cours ou pour préparer le suivant. (C3) Ils l'ont **nettement moins utilisée dans une perspective de reddition de compte (C4) ou de partage des leçons** avec de tierces parties (C5). Il ressort par ailleurs que les **capacités des ANG en évaluation ont progressé, mais modestement**, quand bien même la responsabilité de l'évaluation leur avait été donnée. (C6)

L'étude montre que **la qualité des rapports d'évaluation pose problème**. Même si elle est bonne dans l'ensemble pour une grande majorité des programmes d'universités et d'institutions scientifiques, elle est faible pour la moitié des programmes des ONG et des autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité. (C7) La **crédibilité** des évaluations visant à rendre compte des résultats est par ailleurs **affectée** par le fait que les ANG en sont les commanditaires et que les garanties d'indépendance ne sont pas toujours suffisantes. (C4)

Néanmoins, malgré les faiblesses susmentionnées dans la politique et les rapports d'évaluation, les recommandations de celles-ci ont été considérées comme **utiles** par la plupart des ANG. Les gestionnaires de la DGD ont par ailleurs largement apprécié la complémentarité et la valeur ajoutée des évaluations par rapport au rapportage annuel par les ANG. (C10)

Enfin, l'examen des conclusions des rapports montre que parmi les points positifs fréquemment soulignés, on retiendra le fait que **les programmes répondaient à des besoins clairement identifiés et étaient cohérents avec les stratégies nationales**. De même les rapports affirment souvent que **les activités/réalisations attendues ont été atteintes** dans la plupart des cas. Ils sont par contre plus critiques sur la **conception formelle des stratégies qui semble souvent poser problème**, ou encore sur les **systèmes de suivi et d'évaluation rarement performants**. (C8) Les évaluations manquaient fréquemment d'étudier les résultats ('outcomes') au niveau des populations cibles. La moitié des rapports n'examinait en fait pas la durabilité de l'action, ni l'impact (potentiel). Il n'est donc **pas possible de déterminer dans quelle mesure la coopération non gouvernementale de la DGD dans son ensemble a effectivement contribué au développement**. (C9)

Recommandations

La DGD est actuellement confrontée à des défis importants en tant que bailleur de fonds gouvernemental, et ce d'autant plus que le pays est confronté à une crise budgétaire majeure. Elle a l'obligation de rendre compte de l'utilisation efficace et efficiente de l'argent public qu'elle consacre à la coopération au développement. Elle doit également améliorer l'efficacité de son aide dans l'esprit des engagements internationaux des dernières années, tel que le Partenariat de Busan (2011) ; la société civile a d'ailleurs également souscrit à ce dernier.

Il est ainsi recommandé à la DGD de **définir une politique d'évaluation claire** de la coopération non gouvernementale. Celle-ci devrait distinguer clairement entre trois types d'objectifs potentiels pour l'évaluation : (i) rendre compte ; (ii) aider à la prise de décision ; et (iii) capitaliser les connaissances et partager les leçons. La DGD devrait définir son approche pour chacun des types d'objectifs retenus. Elle devrait sur cette base déterminer les modalités, les dispositifs et un cadre réglementaire adéquats, ainsi qu'une allocation appropriée des responsabilités et des ressources. (R1)

Par rapport au premier type d'objectif (**rendre compte**), la DGD devra en tous les cas élaborer une approche cohérente pour l'évaluation des résultats dans une perspective de gestion axée sur les résultats, comme exigé par la toute nouvelle Loi relative à la Coopération au Développement. Il faudra pour cela s'assurer que le dispositif soit représentatif, consolidable et crédible ; il devrait aussi se baser dans la mesure du possible sur les systèmes de suivi et d'évaluation des ANG. (R2)

Concernant le second type d'objectif (**aide à la prise de décision**), il est recommandé que la **DGD** puisse retrouver, d'une façon ou d'une autre, la capacité de commanditer des évaluations sur des sujets importants pour sa propre prise de décision. Ceci n'est en effet (pratiquement) plus possible depuis une réorganisation en 2010, alors même qu'un rôle plus stratégique est assigné à la DGD au sein de la coopération belge. (R3) Les **ANG** ont par ailleurs montré leur intérêt pour utiliser l'évaluation comme outil d'aide à la décision ; il est important pour une évaluation ayant ce type d'objectif de leur donner une grande liberté d'action, afin qu'ils puissent au mieux la faire correspondre à leurs besoins. (R4) Pour le troisième type d'objectif (**capitaliser sur les connaissances et partager les leçons**), il serait utile que la DGD et les ANG élaborent un mécanisme pour favoriser un partage actif de l'expérience acquise, en vue d'améliorer l'efficacité de l'aide. (R5)

La DGD et les ANG doivent de plus s'assurer que l'ensemble des **acteurs** aient des **capacités solides en évaluation**, avec notamment des systèmes de planning, suivi et évaluation performants. Il est recommandé à ce propos de concevoir un dispositif fort, combinant (i) des exigences de qualité strictes et (ii) de l'appui à la qualité des évaluations et au renforcement des capacités des acteurs. Des exemples concrets existent, comme aux Pays-Bas ou en France. (R6) La **DGD**, aussi, devrait s'assurer de pouvoir disposer de suffisamment d'expertise en évaluation. (R9)

Enfin, la DGD et les ANG devraient prêter une grande attention aux **facteurs principaux qui ont affecté la qualité des évaluations**. Il s'agit tout d'abord de veiller à ce que la demande soit claire et réaliste, ce qui implique entre autres de clarifier l'objectif prioritaire et d'adapter certaines dispositions du cadre réglementaire actuel (notamment sur le budget). Il faudra également veiller à l'encadrement du processus. Il faudra de plus s'assurer que les évaluations examinent la logique de l'intervention, prennent en compte les critères d'évaluation adéquats et évaluent les résultats, la durabilité et l'impact (potentiel) des interventions sur les destinataires. (R7) Les évaluations devront également avoir lieu au moment le plus utile au regard de l'objectif visé. (R8)

Toutes ces recommandations peuvent être considérées comme urgentes et importantes, dans la mesure où les modalités des subventions de la DGD aux ANG pour le prochain cycle de financement sont en train d'être redéfinies.

1. Introduction

Cette méta-évaluation des programmes des acteurs non gouvernementaux (ANG)² fut commanditée par le Service de l'Évaluation spéciale de la Coopération internationale belge (SES), un service externe rapportant directement au Ministre de la Coopération au Développement (cf. cahier spécial des charges en Annexe 1).

Le présent document est le rapport final de cette étude. Il tient compte des commentaires reçus du SES et du Comité d'accompagnement (CA) sur la version provisoire de ce rapport.

1.1 Objet, objectifs et champ

La figure ci-dessous présente les éléments principaux caractérisant cette étude, qui a trait à la période 2008-2011.

Figure 1 – Objet, objectifs et champ

Objet	Méta-évaluation des programmes des acteurs non-gouvernementaux (ANG) → <i>Évaluation de l'évaluation de ces programmes et synthèse</i>
Objectifs	• Apprécier et tirer les leçons de façon indépendante sur : • Politiques d'évaluation (DGD + ANG) • Qualité, indépendance et crédibilité des rapports • Contenu des rapports d'évaluation • Utilisation des rapports (par DGD + ANG) • Fournir des recommandations sur cette base pour améliorer l'évaluation future des programmes (politiques et rapports)
Champ	• <i>Temporel:</i> 2008-2011 • <i>Thématique:</i> Tous secteurs • <i>Géographique:</i> Tous pays • <i>Institutionnel:</i> Tous les ANG cofinancés et la DGD

Source: ADE sur base des Termes de référence

² La dénomination d' « acteur non gouvernemental » (ANG) utilisée durant la période d'évaluation 2008-2011 a été remplacée depuis lors par celle d'« acteur de la coopération au développement non gouvernementale » (ACNG). A des fins de clarté, nous utilisons de façon uniforme dans ce rapport la dénomination qui était d'application durant la période d'évaluation, soit celle d'ANG.

La présente étude répond ainsi aux deux définitions d'une méta-évaluation données par le Comité d'aide au développement de l'OCDE (cf. Glossaire ci-avant):

Méta-évaluation : « Évaluation conçue comme une synthèse des constatations tirées de plusieurs évaluations. Le terme est également utilisé pour désigner l'évaluation d'une évaluation en vue de juger de sa qualité et/ou d'apprécier la performance des évaluateurs. »

Par '**acteurs non gouvernementaux**' (ANG), il est fait référence dans cette étude aux ONG, universités, institutions scientifiques, pouvoirs locaux, syndicats et d'autres partenaires dont un programme a été cofinancé par la DGD durant la période d'évaluation au titre de la 'coopération bilatérale indirecte'. Ces différents acteurs sont fréquemment catégorisés dans cette évaluation sur base de celle utilisée par la DGD dans son cadre réglementaire et au sein de l'ex-direction D3 'Programmes non gouvernementaux' durant la période étudiée ³:

- **ONG**⁴ *ex-D3.1*
- **Universités et institutions scientifiques** *ex-D3.2*
- « **Autres partenaires** » : syndicats, pouvoirs locaux et autres⁵ *ex-D3.3*

1.2 Structure du rapport

Ce rapport est structuré en quatre grands chapitres :

- une description de la **Méthodologie** appliquée ;
- les **Réponses aux questions d'évaluation**, décrivant les faits, l'analyse et les constats sur base des questions définies dans le cahier des charges, ainsi que des éléments de contexte comme une description du cadre réglementaire et de son historique ;
- les **Conclusions**, incluant également une appréciation générale ;
- les **Recommandations** qui en découlent.

Le **volume d'annexes** fournit par ailleurs des détails sur, entre autres, la méthodologie et les outils d'évaluation, les résultats de l'analyse statistique, les réponses aux trois enquêtes, la liste des personnes rencontrées et la bibliographie. Voir la liste des annexes dans la Table des matières au début de ce rapport.

³ Cette direction D3 a été réformée en 2012 et est à présent organisée sur base géographique. Chaque (sous-)direction géographique couvre maintenant les différents types d'acteurs non gouvernementaux actifs dans la région. L'organisation actuelle est la suivante : Direction D3 « Société civile » composée de la D3.1 « Éducation au développement », la D3.2 « Afrique du Nord et de l'Ouest, Monde arabe », la D3.3 « Afrique centrale et australe » et la D3.4 « Asie et Amérique latine ».

⁴ Le champ de l'évaluation ne couvre que les « ONG-programmes ». Ce choix a été fait par le commanditaire pour garder une cohérence et un focus à l'étude.

⁵ Les « autres partenaires » ayant bénéficié de cofinancement de leur programme durant la période 2008-2011 sont les suivants: trois syndicats (FGTB/IFSI, CGSLB/MSI, CSC/IEOI), des pouvoirs locaux (AVCB, UVCW, Province de Hainaut, VVSG), ainsi que Africalia, APEFE, ATOL, Benelux Afro Center et VVOB.

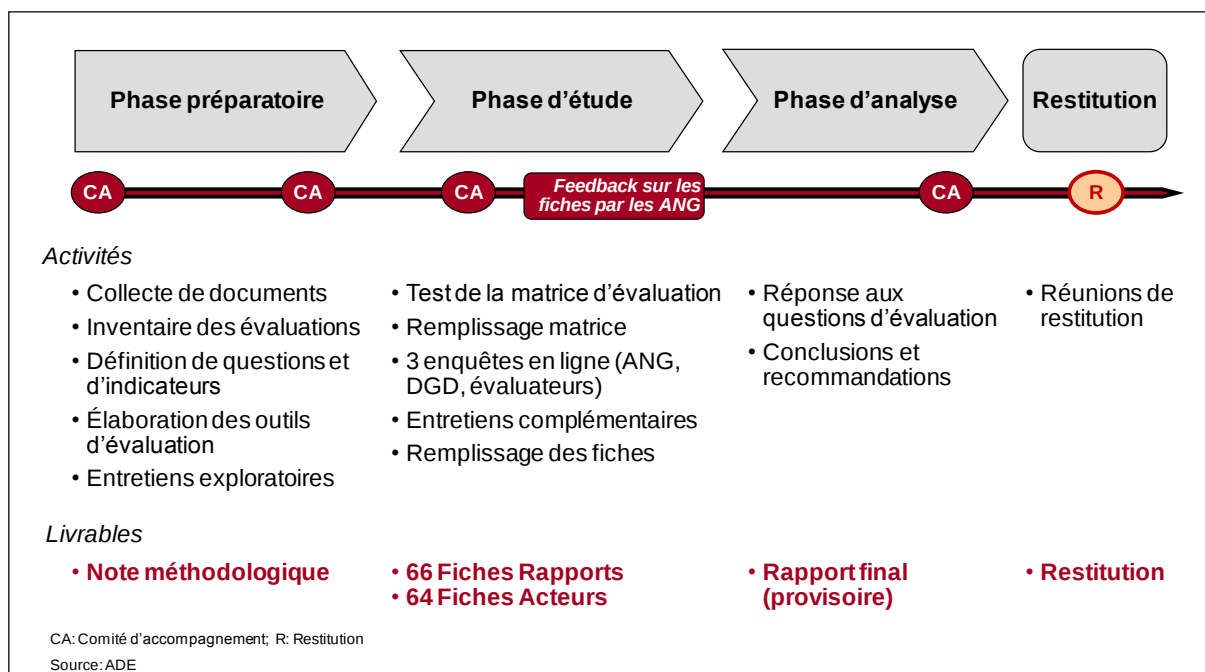
2. Méthodologie

Ce chapitre présente la méthodologie innovante qui a été développée et utilisée pour cet exercice particulier de méta-évaluation. Il décrit en particulier (i) le processus d'évaluation général, (ii) la sélection de l'échantillon, (iii) les questions d'évaluation, (iv) les différents outils utilisés et (v) les défis et limites de l'exercice.

2.1 Processus d'évaluation

Cette étude s'est basée sur un processus composé de plusieurs phases tel que défini par le commanditaire dans le cahier des charges (cf. Annexe 1). Ce processus est présenté dans la figure ci-dessous. Celle-ci indique la séquence des différentes activités réalisées dans le cadre de cette méta-évaluation, ainsi que les livrables intermédiaires et finaux.

Figure 2 - Processus d'évaluation



Ce processus d'évaluation a été géré par le SES, assisté d'un CA dans une perspective participative. Celui-ci comprenait les deux fédérations d'ONG⁶, un représentant de chaque autre acteur non gouvernemental, des gestionnaires de la DGD ainsi qu'un membre du cabinet du Ministre de la Coopération au Développement. Ce comité a été impliqué dans le suivi du travail d'évaluation à chaque phase du processus, comme indiqué dans la figure ci-dessus.

Cette approche participative visait à renforcer la qualité de l'information collectée et à favoriser un apprentissage continu et une appropriation progressive des leçons apprises par chacune des parties prenantes à ce processus, en particulier par la DGD et les ANG. Leur représentation dans le CA a permis de les consulter à chaque étape.

⁶ ACODEV est la fédération francophone et germanophone des associations de coopération au développement ; ngo-federatie (anciennement COPROGRAM) est la fédération flamande des ONG de coopération au développement.

Chacun des ANG a en outre eu l'occasion de commenter la fiche rapport et la fiche acteur qui le concernaient, avant la phase d'analyse globale. La collaboration active entre les différentes parties tout au long du processus a ainsi contribué, outre à la qualité, également à l'utilité et à la crédibilité de cette méta-évaluation.

L'évaluation a démarré fin août 2012 et a été conduite dans un délai de neuf mois comme prévu dans le cahier des charges, si ce n'est que la traduction du rapport a été délivrée plus tard.

2.2 Échantillon

Un échantillon de **66 rapports (de 64 ANG)** a été défini au début de la phase préparatoire par le SES en accord avec l'équipe d'évaluation. Il inclut un rapport par ANG, pour pratiquement tous les ANG dont un programme a été cofinancé durant la période 2008-2011.⁷ L'échantillon visait en particulier à inclure l'évaluation de l'ensemble du programme triennal d'un ANG, pour autant qu'une telle évaluation fût conduite. A défaut, un rapport s'en rapprochant le plus possible fut sélectionné.

L'échantillon et un descriptif de son processus de sélection sont présentés en **Annexe 3**. Il était composé de rapports des types d'acteurs suivants :

- 49 évaluations de programmes d'ONG ;
- 6 évaluations de programmes d'universités et d'institutions scientifiques ;
- 11 évaluations de programmes d'« autres partenaires ».

L'échantillon des rapports peut être considéré comme donnant une image assez juste de l'évaluation auprès de l'ensemble des ANG, dans la mesure où il contient un rapport pour pratiquement chaque ANG. Au niveau des types d'acteurs, l'interprétation des chiffres nécessite davantage de prudence, vu le nombre de cas plus limité, en particulier pour les universités et institutions scientifiques, et pour les « autres partenaires » ; le nombre était plus élevé pour les ONG (cf. ci-dessous).

2.3 Questions d'évaluation

Le questionnement évaluatif a été structuré sur base des sept questions d'évaluation définies dans le cahier des charges. Chacune de ces questions a ensuite été structurée par l'équipe d'évaluation en sous-questions, pourvus d'une série d'indicateurs. Le set de questions d'évaluation est présenté en introduction au chapitre 3 ; les questions complètes avec sous-questions et indicateurs sont-elles fournies en Annexe 4.

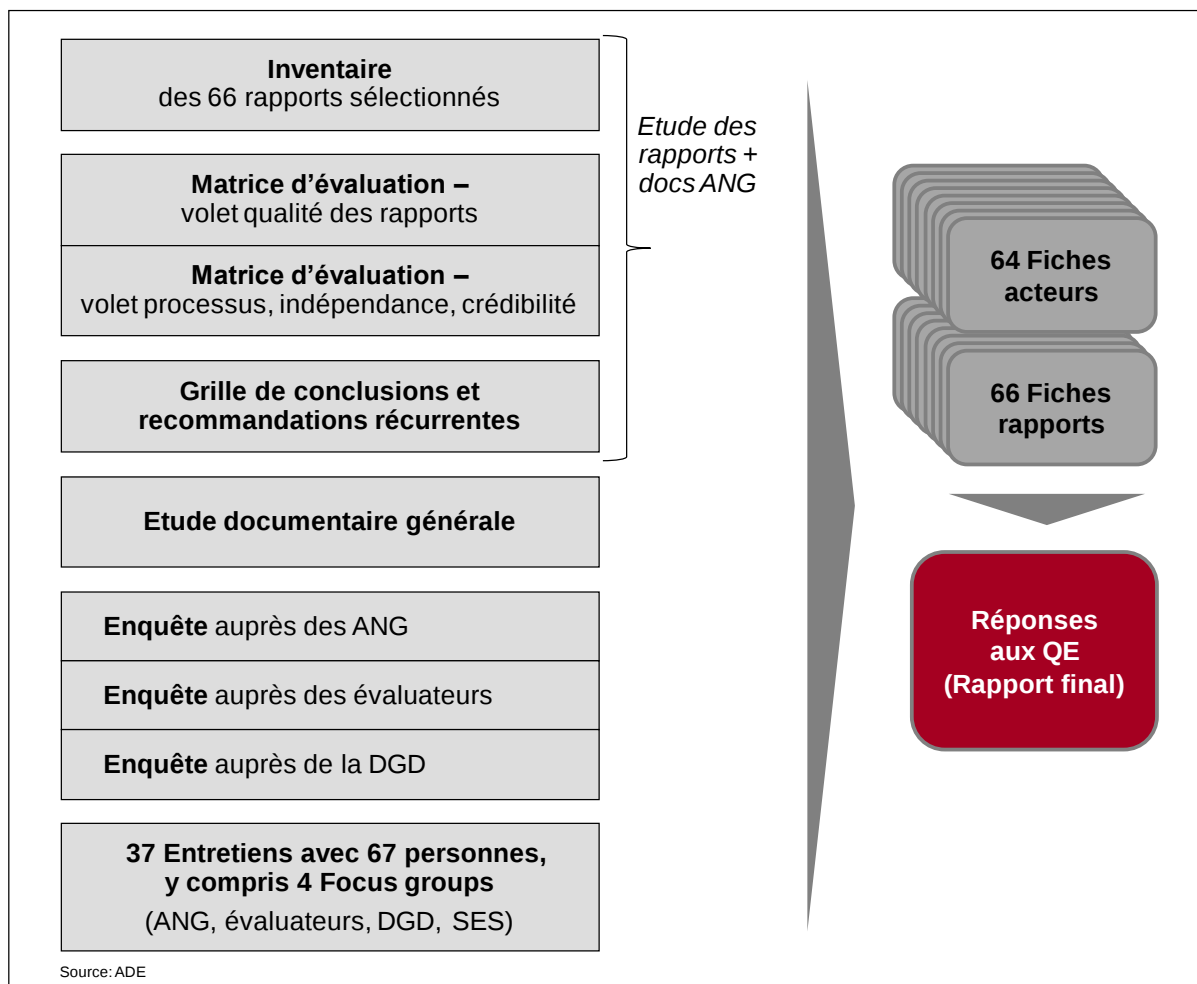
Les constats de cette étude sont présentés dans le chapitre 3 sous la forme de réponses à ces questions d'évaluation.

⁷ 1 rapport d'évaluation a été sélectionné par ANG, à l'exception de la coopération universitaire pour lesquels deux rapports ont été sélectionnés par acteur afin de disposer d'une masse critique : 2 pour les universités flamandes et 2 pour les universités francophones.

2.4 Outils d'évaluation

Une série d'outils d'évaluation ont été définis au début de l'évaluation pour collecter et analyser l'information recherchée. La figure ci-dessous synthétise les principaux outils.

Figure 3 – Outils d'évaluation



La définition de ces outils s'est basée de façon générale sur une revue des pratiques en matière de méta-évaluation, en ce compris des pratiques de la coopération européenne, néerlandaise, française et finlandaise, et de certains types d'acteurs.⁸ Certains d'entre eux ont un caractère innovant, dont la matrice d'évaluation et la grille de conclusions et recommandations récurrentes, comme décrit dans la section 2.5 ci-dessous.

Les outils ont été discutés avec le SES et le CA. La matrice d'évaluation, l'inventaire et la grille de contenu récurrent ont par ailleurs fait l'objet d'un test initial sur base de l'examen de six rapports. Les outils ont ainsi pu être affinés sur cette base avant leur approbation et leur utilisation effective.

⁸ Exemples: Development Researchers' Network, 2008. *Synthesis of the geographical evaluations managed by the Evaluation Unit during the period 1998-2006*, European Commission ; Ministerie van Buitenlandse Zaken (in Nederland), 2012. *Betreft IOB-rapport "De methodische kwaliteit van Programma-evaluaties in het Medefinancieringsstelsel-I (2007 – 2010)"*; Finnish Consulting Group Ltd, 2007. *Meta-Analysis of Development Evaluations in 2006*. Ministry for Foreign Affairs of Finland, Department for Development Policy ; ITECO, Décembre 2010. Antipodes n°191, *Évaluer l'évaluation*.

Le tableau ci-dessous résume l'utilisation de chacun des outils par questions d'évaluation.

Tableau 1 – Outils d'évaluation, par question d'évaluation

Outils d'évaluation	Q1 Politiques	Q2 Qualité	Q3 Indép.	Q4 Crédibilité	Q5 Contenu	Q6 Util. ANG	Q7 Util. DGD
Inventaire des rapports	■	□	□				
Matrice d'évaluation des rapports et pratiques	■	■	■	■		■	
Grille de contenu récurrent des rapports					■		
Etude documentaire générale	■		□			■	
Enquêtes	■	■	■	■		■	■
Entretiens et focus groups	■	■	■	■		■	■

Source : ADE



Source importante



Source secondaire

Chaque outil et son utilisation sont présentés et décrits en détail dans une annexe, accompagnés le cas échéant de l'information recueillie par cet outil et de l'analyse des données:

- Annexe 5 - Matrice d'évaluation (et résultats)
- Annexe 6 - Fiche rapport (et résultats)
- Annexe 7 - Fiche acteur (et résultats)
- Annexe 8 - Enquête ANG (et résultats)
- Annexe 9 - Enquête évaluateurs (et résultats)
- Annexe 10 - Enquête DGD (et résultats)
- Annexe 11 - Grille de conclusions et recommandations récurrentes
- Annexe 13 - Liste des personnes rencontrées
- Annexe 14 - Bibliographie

2.5 Défis et limites

Cette méta-évaluation était confrontée à un certain nombre de défis. Ceux-ci sont liés en particulier (i) à l'absence d'un cadre de référence formel, (ii) à la rigueur de l'approche méthodologique, (iii) à la fiabilité des données, (iv) à l'approche participative pour favoriser l'utilité de cette méta-évaluation, et (v) au calendrier de l'évaluation.

Absence d'un cadre de référence formel

Comme il ressort de l'analyse, la DGD n'a pas défini de réelle politique d'évaluation ni défini d'objectifs et de directives claires en matière d'évaluation. Il n'y avait ainsi **pas de cadre de référence formel pour structurer la méthodologie**. L'équipe d'évaluation a ainsi développé un cadre de référence sur l'évaluation propre au présent exercice sur base des politiques et pratiques d'autres bailleurs et acteurs (cf. section 2.4 ci-dessus), et de sa propre expérience en évaluation. Celui-ci a été discuté avec le SES et le CA pour le vérifier et en assurer l'appropriation par les différentes parties à cette méta-évaluation.

Rigueur de l'approche méthodologique

Développement d'une méthodologie rigoureuse et innovante: La conception de cette « méta-évaluation » et des outils d'évaluation a constitué un défi méthodologique. Il s'agit en effet d'un type d'étude qui requiert une rigueur méthodologique importante, mais pour lequel il y a relativement peu de référents ou d'antécédents, en particulier dans le domaine de la coopération internationale. La méthodologie a ainsi été développée avec une grande attention. Elle est en ce sens unique et innovante à plus d'un titre. Tout en tirant les leçons d'exercices antérieurs, l'équipe d'évaluation a en effet créé ou développé davantage des outils d'évaluation permettant de répondre aux besoins spécifiques de cette étude de façon innovante, tels que la **matrice d'évaluation** (avec un volet sur la qualité des rapports et un autre sur les politiques et pratiques des acteurs en matière d'évaluation), les **fiches rapports** et **fiches acteurs**, ou la **grille de catégorisation des conclusions et recommandations récurrentes**.

Fiabilité des données

Objectivité et harmonisation : Il est important dans un tel exercice de s'assurer de porter un jugement le plus objectif et homogène possible sur l'ensemble des rapports et pratiques d'évaluation. Une attention particulière a été donnée à cette dimension lors de la conception des outils. Ceci a résulté entre autres en une matrice d'évaluation qui consiste en une série d'indicateurs pour lesquels des valeurs de jugement harmonisées ont été définies. Les trois enquêtes (auprès des ANG, des évaluateurs et des gestionnaires de la DGD) ont par ailleurs permis d'avoir des données systématiques et homogènes sur certains indicateurs. En outre, l'équipe d'évaluation a rédigé des directives précises pour avoir des appréciations les plus homogènes possibles par les membres de l'équipe et par les différents types d'acteurs. L'étude des rapports et des documents des ANG a par ailleurs été cadrée étroitement par le chef d'équipe, afin de pouvoir répondre dans les meilleurs délais à toute question d'un membre de l'équipe. Ceci inclut entre autres que les experts ont fait la plupart du travail d'analyse des rapports dans la même pièce ou physiquement proches, afin de favoriser l'interaction et le feedback immédiats. Ces dispositions ont montré leur utilité. Enfin, la variété d'outils a permis de recueillir des données tant qualitatives que quantitatives, et tant de première que de seconde main. Cela a permis également de croiser ('triangler') l'information des différentes sources et a ainsi contribué à l'objectivation du jugement final. Néanmoins, en dépit de tous ces efforts pour objectiver et harmoniser les appréciations, les jugements portés dans cette méta-évaluation garderont toujours une part importante de subjectivité, par exemple de par la définition-même des indicateurs et de leurs valeurs de jugement (faite toutefois en concertation avec le CA). Il est également important de

préserver l'esprit critique et la capacité de jugement d'ensemble de l'équipe d'évaluation dans l'utilisation des outils.

« Représentativité » et ampleur de l'exercice : L'« échantillon » pour cette méta-évaluation, défini en concertation avec le SES et le CA, **couvre pratiquement tous les acteurs non gouvernementaux** dont un programme triennal a été cofinancé par la DGD durant la période 2008-2010 ou 2009-2011 (voir Annexe 3). Il est relativement inhabituel d'obtenir une telle couverture du champ dans une méta-évaluation. Ceci implique néanmoins un défi particulier quant à l'**ampleur de l'exercice**. L'information a ainsi été collectée auprès d'une centaine de personnes (64 ANG, 23 gestionnaires de la DGD, plus de 50 bureaux d'évaluation et évaluateurs indépendants contactés, etc.). Ceci a aussi entraîné d'examiner en détail des centaines de documents (66 rapports d'évaluation avec leurs annexes et cahier des charges, les documents sur les pratiques internes des 64 ANG, de la documentation générale, etc.). C'est pour capturer, croiser et analyser l'information de ces différentes sources sur cette multitude de cas qu'il a fallu créer une série de près d'une douzaine d'outils d'évaluation.

Qualité des informations recueillies : La qualité de l'analyse dépend également, comme souvent, de la qualité des informations et contributions fournies par les différents acteurs. La collaboration du SES, de la DGD et des acteurs non gouvernementaux a ainsi été particulièrement importante. L'exercice a par ailleurs prêté une attention particulière à la participation des différents acteurs au processus en vue de contribuer à la qualité de la base factuelle pour la méta-évaluation.

Absence de conflit d'intérêt : L'équipe d'évaluation étant amenée à se prononcer sur la qualité d'autres évaluations, il était important de veiller à éviter tout conflit d'intérêt ou même de perception en la matière. ADE n'a ainsi conduit aucune des évaluations sélectionnées dans l'échantillon, ni d'autres évaluations de programmes d'ANG belges durant la période couverte ou depuis lors. L'équipe d'évaluation a par ailleurs appliqué les plus hauts standards d'éthique et d'indépendance durant le présent exercice, afin d'éviter tout biais personnel ou toute perspective corporatiste. Le SES et le CA ont par ailleurs encadré l'exercice, visant entre autres à vérifier la fiabilité des informations et la qualité de l'argumentation.

Il reste toutefois certaines **limites spécifiques** en termes de représentativité et de validité de l'information recueillie:

- L'équipe d'évaluation n'a pas eu accès aux **offres** des bureaux d'évaluation ou consultants indépendants répondant au cahier des charges des évaluations des programmes des ANG, étant donné le caractère sensible et confidentiel de la chose. La méta-évaluation ne dispose en conséquence pas d'information de première main sur, par exemple, les processus d'appel d'offres, le budget des évaluations, le profil des évaluateurs ou la méthodologie initialement proposée par l'équipe d'évaluation. Des informations sur le sujet ont toutefois été collectées dans une certaine mesure au travers des enquêtes et des entretiens avec les ANG et les évaluateurs.
- Les entretiens avec des **évaluateurs** ont été réalisés principalement avec des consultants de bureaux d'évaluation, sélectionnés en fonction du nombre d'évaluation conduites par ces bureaux (afin de maximiser la couverture du nombre de rapports pour le nombre d'entretiens prévus avec des évaluateurs). Une enquête spécifique a toutefois permis de couvrir les équipes d'évaluation de près de la moitié des rapports. Ceci dit, le questionnaire étant disponible en français et en néerlandais, cela a sans doute freiné des évaluateurs ayant une connaissance insuffisante de ces deux langues (il y a également des rapports d'évaluation en anglais et quelques rapports en espagnol et en portugais). La traduction vers ces autres langues n'a pas été réalisée pour des raisons de coût (les coûts de la production de chacune des trois enquêtes en deux langues étant déjà élevés). En outre, l'accès à internet a pu être une difficulté pour la participation de certains consultants locaux à l'enquête en ligne. Il a été tenu compte de ces biais potentiels dans l'analyse des données.

- Il n'a par ailleurs pas été possible dans cette méta-évaluation de recueillir de l'information auprès des **parties prenantes directes des projets** (notamment les acteurs locaux dans les pays d'intervention), par exemple concernant leur participation aux évaluations ou leurs vues sur la qualité et l'utilité de ces évaluations. Ceci était lié au à la complexité et au coût additionnels important pour ce faire.
- L'équipe a pu rencontrer la plupart des personnes qu'elle désirait rencontrer. Elle a également obtenu une participation très importante aux trois enquêtes. Il peut toutefois exister un biais de ce point de vue, comme toujours, dans la mesure où les personnes les plus enclines à fournir de l'information peuvent être celles qui ont les avis les plus tranchés.

Utilité de la méta-évaluation

Approche participative : Le processus de la méta-évaluation a été pensé dans une perspective participative, incluant la DGD, les ANG ainsi que le cabinet du Ministre de la Coopération au Développement (cf. section 2.1 ci-dessus). Un défi de cette approche participative, outre celui d'une complexité accrue du processus, est celui des attentes différentes des nombreuses parties prenantes. Ceci a été géré en collaboration avec le SES, en visant à faire bénéficier les différentes parties des leçons de cet exercice d'évaluation, tout en restant centrés sur l'objectif principal qui est de rendre compte de la politique et des pratiques de la DGD et des ANG en matière d'évaluation (cf. TdR).

Calendrier

Un calendrier serré : Le calendrier de cette méta-évaluation, tel que défini dans le cahier des charges, était serré compte tenu du défi méthodologique de l'évaluation et de l'ampleur du travail à effectuer (cf. supra). Outre le défi en termes de ressources humaines, cela a par exemple imposé d'utiliser les outils en parallèle, sans pouvoir affiner les questions d'un outil sur base de l'analyse des résultats des autres outils. Des solutions et de la flexibilité ont toutefois été trouvées pour garantir la qualité du rapport.

3. Réponses aux Questions d'évaluation

Les **constats** de l'évaluation sont présentés dans ce chapitre, sous la forme de réponses aux questions d'évaluation. Ils sont structurés sur base du cadre d'analyse global défini pour l'évaluation, qui consiste en sept **questions d'évaluation** (Q). Celles-ci se composent chacune de **sous-questions** et d'une série d'indicateurs (I). Ce set de questions d'évaluation structurées complète et détaille celui proposé dans le cahier des charges (cf. annexe 1).

Un aperçu des questions d'évaluation est présenté dans le tableau ci-dessous (les questions détaillées avec leurs sous-questions et indicateurs sont présentées en Annexe 4). La réponse à chacune de ces questions est présentée par la suite.

Tableau 2 – Aperçu des questions d'évaluation

Q1	Politiques	Quelles exigences la DGD a-t-elle définies en matière de suivi et d'évaluation des programmes cofinancés ? Comment sont-elles suivies par les ANG ? Permettent-elles d'avoir des informations essentielles sur les résultats obtenus ? Comment les directives et leur application peuvent-elles être améliorées ?
Q2	Qualité	Les évaluations répondent-elles aux besoins et sont-elles de bonne qualité ? Quelle a été l'incidence du processus d'évaluation en matière de qualité de l'évaluation et d'apprentissage pour les parties prenantes ? Comment améliorer la qualité ?
Q3	Indépendance	Dans quelle mesure les évaluations sont-elles indépendantes ?
Q4	Crédibilité	Quelle est la crédibilité des évaluations ?
Q5	Contenu	Y a-t-il des constats, des conclusions ou des recommandations qui reviennent toujours dans les rapports des acteurs non gouvernementaux ? Quels sont les principaux points positifs ou négatifs de l'ensemble des rapports d'évaluation ?
Q6	Utilisation par les ANG	Comment les ANG utilisent-ils les évaluations ?
Q7	Utilisation par la DGD	Comment la DGD utilise-t-elle les évaluations ?

Source : ADE sur base des TdR

3.1 Politiques et pratiques de la DGD et des ANG – Q1

Quelles exigences la DGD a-t-elle définies en matière de suivi et d'évaluation des programmes cofinancés ? Comment sont-elles suivies par les ANG ? Permettent-elles d'avoir des informations essentielles sur les résultats obtenus ? Comment les directives et leur application peuvent-elles être améliorées ?

La réponse à cette question vise tout d'abord à déterminer le cadre réglementaire fixé par la DGD en matière d'évaluation des programmes des ANG. Elle étudie dans quelle mesure il était clair et détaillé, en ce compris en termes de reddition de compte sur les résultats. Elle examine également son degré d'harmonisation pour les différents types d'ANG. Elle porte ensuite le respect de ce cadre réglementaire par les ANG. La question examine enfin les capacités en évaluation des ANG et leurs politiques et pratiques internes en la matière. Les recommandations relatives à ces différents points sont formulées dans le dernier chapitre de ce rapport.

Q1 - Politiques et pratiques de la DGD et des ANG

Résumé de la réponse

Pour la période 2008-2011, il n'y avait pas à la DGD de véritable politique d'évaluation fixant des objectifs, définissant des modalités et prévoyant les ressources nécessaires. Celle-ci s'est largement limitée à transférer la responsabilité de l'évaluation aux acteurs. Les directives en matière d'évaluation consistaient en des références plus ou moins explicites dans une série de documents de types divers et variés. Elles laissaient aux ANG une grande liberté en la matière, avec un rôle souvent limité voire inexistant de la DGD, en particulier pour les ONG. Les gestionnaires de la DGD ont toutefois souvent dû donner des clarifications et ont en outre donné des instructions personnelles additionnelles, sur base de leur sensibilité à l'évaluation. Les instructions concernaient en particulier les objectifs de l'évaluation (73%), le moment de l'évaluation (73%), les critères d'évaluation (73%), l'objet de l'évaluation (60%), l'envoi du rapport à la DGD (57%) et des mesures pour favoriser l'indépendance de l'évaluation (53%).

Les ANG ont largement apprécié la flexibilité offerte et ont globalement respecté le cadre réglementaire. Il y a des indications que durant la période, la capacité d'évaluation en matière d'évaluation a été renforcée dans une majorité d'ANG: présence de personnes capables de gérer des évaluations, participation à des formations à l'évaluation, etc. Cependant, près d'un tiers des ANG a connu une augmentation limitée, et un tiers des ONG n'a pas réalisé d'évaluation à proprement parler pour les activités cofinancées. Par ailleurs, il n'y a pas à ce stade, pour la plupart d'entre eux, de véritable politique d'évaluation explicite. Il y a, ceci dit, des indications que la DGD a contribué en partie au renforcement (modeste) des capacités en évaluation de la moitié des ONG, par les dispositions de son cadre réglementaire, même si peu exigeantes, et par les formations organisées par les fédérations d'ONG qu'elle finançait. Ceci était toutefois bien moins le cas pour les autres types d'acteurs.

3.1.1 Politique et pratiques de la DGD (Q1.1 et Q1.2)

Politique d'évaluation

La DGD se basait pour l'évaluation des programmes des acteurs non gouvernementaux durant la période 2008-2011 sur le transfert à leur niveau de la responsabilité en matière d'évaluation. La réforme de la coopération au développement à la fin des années 1990 visait une responsabilisation et une professionnalisation de ces acteurs en matière d'évaluation. L'administration a donné sur cette base aux acteurs (du Nord) la charge de commanditer et de gérer l'évaluation dans le cadre de leur programme. Cependant, **la DGD n'avait pas défini de politique d'évaluation en tant que telle** durant cette période, alors que la réforme prévoyait que l'administration devînt responsable « de la préparation et de l'évaluation de la politique ».⁹

Cadre réglementaire

Le cadre réglementaire était relativement distinct selon le type d'acteur : très peu directif pour les ONG, un peu plus pour les universités et institutions scientifiques, et davantage pour les autres partenaires. Il restait toutefois relativement peu directif dans l'ensemble. Les détails du cadre réglementaire sont fournis en Annexe 2 ; les principaux éléments en sont résumés dans ce qui suit.

Pour les ONG, une grande liberté leur était donnée en matière d'évaluation, pratiquement sans implication de la DGD.¹⁰ Le cadre réglementaire de la DGD laissait à chaque ONG le soin de définir sa propre approche pour l'évaluation de son programme. Celle-ci était prise en compte dans l'appréciation par la DGD de la proposition de programme de l'ONG. Le cadre réglementaire de la DGD ne donnait ainsi pas de directives particulières concernant les types de rapports attendus, les méthodologies à suivre, la nécessité ou non de rapporter sur l'atteinte des résultats attendus, l'implication de la DGD, l'envoi des rapports, etc. Il référait implicitement à de l'évaluation externe, mais sans préciser ce qui était entendu par là. Il ne donnait pas non plus d'indication sur l'ampleur du budget à prévoir pour les activités d'évaluation, si ce n'est que cela dit s'inscrire dans le total des frais de gestion¹¹ et que les honoraires de consultance ne pouvaient dépasser 500€ par jour de prestation. La DGD avait en effet rompu en 2006 avec sa politique du « 1% » (1% du budget du programme doit être consacré à de l'évaluation), sans que les raisons de cet arrêt fussent explicitées.¹²

⁹ Sénat de Belgique, 1999. *Projet de loi relative à la coopération internationale belge, procédure d'évocation*, p2.

¹⁰ Le cadre réglementaire pour les ONG était constitué en particulier d'un arrêté royal de 2006 et d'un arrêté ministériel de 2007 relatifs au cofinancement de programmes d'ONG, ainsi que de directives sur le schéma de présentation d'un programme par les ONG (2007), sur l'appréciation des programmes par la DGD (2007) et sur les rapports narratifs des programmes (2008). Voir détails en Annexe 2.

¹¹ Le taux de frais de gestion en pourcentage du budget total du programme était défini individuellement pour chaque ONG en fonction d'une formule complexe et se situait typiquement entre 5% et 9%.

¹² Des explications sur la suppression du « 1% » n'ont pu être trouvées dans des sources écrites. Quelques interlocuteurs actifs à cette période-là ont mentionné que la suppression faisait suite à une demande du secteur des ONG, vu les progrès accomplis grâce entre autres au « 1% » et certains effets pervers de cette mesure. Pour des détails sur le « 1% », voir l'étude d'évaluation de cette mesure réalisée en 2005 (cf. bibliographie en Annexe 4).

Pour les partenaires universitaires, la DGD leur laissait également une grande liberté en matière d'évaluation, tout en restant quelque peu impliquée.¹³ La DGD dit en effet être représentée dans un **comité de concertation** en charge du suivi et de l'évaluation (sans autre spécification sur le rôle d'un tel comité). Le cadre réglementaire prévoyait par ailleurs qu'au **minimum 1%** du budget du programme soit consacré à de l'évaluation (sans davantage de spécifications) ; cette disposition de 1997 est restée valable jusqu'à aujourd'hui. Plusieurs interlocuteurs ont souligné que la DGD partait d'une relation de confiance avec les partenaires universitaires (et scientifiques) en matière d'évaluation, sur base de leur expérience en la matière. Celle-ci est liée à la culture d'évaluation spécifique du monde académique et scientifique, et au fait que ces partenaires considèrent souvent comme naturel de rendre compte de l'utilisation du financement public dont ils ont pu bénéficier, et dont ils dépendent largement.

Pour les institutions scientifiques, la DGD avait ainsi également un cadre réglementaire peu directif en matière d'évaluation.¹⁴ Celui-ci prévoyait pour l'Institut de médecine tropicale (IMT) et pour le Musée royal de l'Afrique Centrale (MRAC) que la responsabilité des évaluations fussent données à un **comité de pilotage**, où la DGD fût représentée (sans autre spécification sur son rôle). Dans le cas du MRAC, les spécifications du protocole de coopération relatives à l'évaluation se limitaient au fait que le comité de pilotage déterminait « les termes de référence, le contenu et le budget de l'évaluation ainsi que la manière dont cela s'organisera » et que le budget pour ces évaluations représentera « au moins 1% du budget total du programme et ne dépassera pas les 5%. » Ces deux points n'étaient pas mentionnés dans la convention de coopération avec l'IMT, qui mentionnait par contre que « les activités des programmes de six ans auront au moins toutes une fois fait une évaluation ».

Pour les autres partenaires¹⁵, le cadre réglementaire était plus précis, tout en restant relativement peu directif sur le fond.¹⁶ Le cadre réglementaire d'application durant la période 2008-2011 précisait que les évaluations dussent être « **engagées en vue d'une amélioration prospective** » ; le rôle de faire le point sur l'atteinte des résultats était lui dévolu au rapportage annuel par les acteurs, sur base de la *base line* et des indicateurs objectivement vérifiables définis dans le cadre de leur programme global de cofinancement.¹⁷ Le cadre réglementaire mentionnait qu'il pouvait y avoir **plusieurs types d'évaluation** (ex ante, à mi-parcours, ex post, etc.) et qu'elles pouvaient être

¹³ Le cadre réglementaire pour les partenaires universitaires était constitué en particulier d'une convention générale et de conventions spécifiques datant de 1997, complété dans certains cas par des conventions ultérieures (comme en 2002 pour les 'Actions-Nord'). Voir détails en Annexe 2.

¹⁴ Le cadre réglementaire pour les institutions scientifiques était constitué d'un accord pour chacun d'eux. Il s'agissait notamment (pour les rapports repris dans notre échantillon) de la Convention de coopération avec l'Institut de médecine tropicale (2008) et du Protocole de coopération avec le Musée Royal de l'Afrique Centrale (2007). Voir détails en Annexe 2.

¹⁵ Pour rappel, les « autres partenaires » ayant bénéficié de cofinancement de leur programme durant la période 2008-2011 sont les suivants: trois syndicats (FGTB/IFSI, CGSLB/MSI, CSC/IEOI), des pouvoirs locaux (AVCB, UVCW, Province de Hainaut, VVSG), ainsi que Africalia, APEFE, ATOL, Benelux Afro Center et VVOB.

¹⁶ Le cadre réglementaire pour les « autres partenaires » était constitué d'un arrêté royal de cofinancement de programme pour chacun d'eux, ainsi que de « documents de référence » standards datant de 2007. Les directives sur l'évaluation proviennent en particulier du document sur les 'Normes financières et dispositions administratives des programmes'. Une version revue de ces documents a été approuvée en 2011 et est d'application à partir du 1^{er} janvier 2012 (soit après la période d'évaluation 2008-2011). Voir détails en Annexe 2.

¹⁷ « Le Rapport Annuel (RA n) du Plan d'Action annuel (PA n) présente de manière concrète les résultats et les activités qui les ont produits durant l'année de référence n concernée. Il fait le point de la situation par rapport aux objectifs et établit une comparaison entre les résultats atteints et ceux qui étaient prévus au PA en se référant aux indicateurs (IOV) préalablement élaborés tant qualitativement que quantitativement. »

internes, externes ou conjointes. Il fallait au minimum une « évaluation externe indépendante » pour chaque programme, voire également une « évaluation à mi-parcours externe indépendante » pour les programmes de plus de quatre ans. Les termes de référence dient en être approuvés par un **comité de concertation**, incluant la DGD, et le choix des consultants externes dit être justifié. Une **synthèse** des rapports internes et externes de maximum 5 pages, ainsi que les rapports de mission, dient par ailleurs être envoyés à la DGD. Il n'y avait néanmoins pas de directives particulières concernant les méthodologies à suivre, l'implication de la DGD au-delà de l'approbation des termes de référence par le comité de concertation, ce qui était entendu par une évaluation « externe » et « indépendante », etc. Il ne donnait pas non plus d'indication sur l'ampleur du budget à prévoir pour les activités d'évaluation, mais précisait que les frais d'évaluation étaient considérés comme des coûts de gestion et donnait une limite d'honoraires de 500€ par jour de prestation.

Instructions additionnelles des gestionnaires de la DGD

Plus de la moitié des gestionnaires de la DGD a donné des instructions personnelles additionnelles pour clarifier ou pour compléter le cadre réglementaire, ce qui a parfois mené à une certaine confusion.

- 56% des gestionnaires de la DGD en moyenne ont affirmé en réponse à l'enquête qu'ils ont donné des instructions additionnelles au cadre réglementaire (26%) ou du moins des clarifications (30%). Ces commentaires étaient souvent donnés dans le cadre de l'appréciation par la DGD de leur proposition de programme et/ou du dialogue de politique de la DGD avec les ANG.
- **Les instructions concernaient en particulier les objectifs de l'évaluation (73%), le moment de l'évaluation (73%), les critères d'évaluation (73%), l'objet de l'évaluation¹⁸ (60%), l'envoi du rapport à la DGD (57%) et des mesures pour favoriser l'indépendance de l'évaluation (53%).**
- Les points du cadre réglementaire qui étaient considérés par les gestionnaires de la DGD comme pas ou peu clairs ou précis étaient les mesures favorisant la qualité et la crédibilité de l'évaluation (25%), l'objet de l'évaluation (19%), les critères d'évaluation (13%) et le moment de conduire l'évaluation (13%). De nombreux autres gestionnaires ont néanmoins répondu à l'enquête qu'il n'y avait en fait « pas d'exigences formelles » (souvent autour de 25% pour les différents critères). La combinaison de ces réponses (exigences formelles pas claires, peu claires ou inexistantes) peut ainsi expliquer pourquoi plus de la moitié des gestionnaires de la DGD a donné des instructions personnelles additionnelles.
- Plusieurs ANG ont indiqué que de telles instructions additionnelles étaient souvent fonction du gestionnaire en charge de leur dossier, et qu'une succession de gestionnaires entraînait fréquemment des attentes différentes. Ceci a été confirmé par des gestionnaires de la DGD, précisant que pour eux non plus il n'était pas clair ce que la DGD attendait de façon générale des évaluations.

¹⁸ Programme triennal cofinancé, activités spécifiques, volets Nord et/ou Sud, etc.

Compréhension et respect par les ANG du cadre réglementaire

Relativement peu d'ANG ont rencontré des problèmes dans la compréhension des exigences de la DGD. La majorité ont apprécié la flexibilité offerte par le cadre réglementaire globalement peu directif. Les 'autres partenaires', qui étaient soumis à des directives plus détaillées, font toutefois exception.

- Relativement peu d'ANG ont indiqué dans l'enquête avoir rencontré des problèmes dans la compréhension du cadre réglementaire de la DGD (9/63 ou 14% au total), ou avoir manqué de flexibilité (5/57 ou 12%). Les 'autres partenaires' forment toutefois une exception notoire, la moitié d'entre eux ayant rapporté des difficultés et un manque de flexibilité. Ceci s'explique probablement par les directives plus détaillées pour eux que pour les ONG, universités ou institutions scientifiques (cf. supra). Il a également été observé dans plusieurs évaluations que, via la participation de la DGD (et dans certains cas du SES) au sein du comité de concertation, des instructions ont été données qui allaient au-delà des dispositions réglementaires (peu directives).¹⁹
- Du côté des gestionnaires de la DGD, également peu d'entre eux ont indiqué avoir rencontré des problèmes de compréhension du cadre réglementaire par les ANG (11% en moyenne sur tous les critères), y compris -et étonnamment- par les 'autres partenaires'.
- De nombreuses ONG ont indiqué lors d'entretiens et de focus groups avoir apprécié la liberté presque totale qui leur avait été donnée en matière d'évaluation. À l'époque du « 1% », l'évaluation était perçue par nombre d'entre elles comme une « évaluation-contrôle » pour la DGD. Cette perception a eu des effets contre-productifs comme un manque de collaboration à l'exercice ou la volonté de s'assurer que les évaluations ne débouchent sur des messages trop critiques.²⁰ Durant la période 2008-2011, au contraire, les ANG ont pu utiliser l'évaluation en fonction de leurs besoins individuels, ce qu'elles ont largement fait dans une optique d'aide à leur décision (voir Q6). Ceci a certainement contribué à élargir encore davantage leur prise de conscience sur l'utilité de l'outil « évaluation ».

¹⁹ Ceci était par exemple lié à la rédaction des termes de référence, à la définition de l'objectif de l'évaluation, à la définition du champ, à la procédure d'appel d'offre (y compris la sélection des évaluateurs), etc. Des « autres partenaires » ont rapporté à titre d'exemple avoir été confronté à des exigences relatives à un canevas formel pour les TdR, un objectif de reddition de compte plus que d'aide à la décision pour l'ANG, une extension du champ à des actions considérées comme moins importantes à évaluer par l'ANG, ou une formalisation avancée de la procédure d'appel d'offre et de la méthode de sélection des évaluateurs.

²⁰ La politique du « 1% » a toutefois contribué, malgré certaines résistances, à insuffler la culture de l'apprentissage et de l'évaluation à de nombreux acteurs. Voir Iram, 2005. *Évaluation de l'utilisation du 1% du programme quinquennal des ONG consacré à l'évaluation, rapport de synthèse*, Service Public Fédéral Affaires Étrangères, Commerce extérieur et coopération au développement, Direction Générale de la Coopération au Développement (DGCD).

Encadré 1 – Coûts de gestion et Limite du niveau d'honoraires

Un problème soulevé par de nombreux ONG et 'autres partenaires' lors des focus groups, entretiens et réponses à l'enquête est la façon dont ils pouvaient budgétiser les frais d'évaluation.²¹ Le cadre réglementaire considérait ces frais comme des **coûts de gestion** du programme. Ceux-ci étant plafonnés, le budget pour l'évaluation a fréquemment dû laisser place à d'autres priorités pour l'ANG, telles que des frais de personnel ou de formation, et ce surtout en fin de programme. Ceci a eu un effet significatif sur le budget des évaluations des ONG, d'après de nombreux interlocuteurs. L'effet fut moindre pour les 'autres partenaires', qui ont pu systématiquement obtenir une dérogation en fin de programme.²²

De nombreux ANG et évaluateurs ont par ailleurs indiqué que la limitation des honoraires à **€500 par jour** était peu claire, mais également irréaliste et même contre-productive. Les ANG affirment avoir reçu des interprétations différentes en fonction des gestionnaires de la DGD, par exemple si la limite est d'application s'ils reçoivent plus de trois offres, ou si elle peut être considérée comme une moyenne. De nombreux ANG ont également mentionné que ce taux ne tenait pas compte des réalités du marché et les a empêchés d'avoir accès à des experts reconnus ou à des bureaux d'étude. Ceci même en considérant que de nombreux évaluateurs (du Nord) ont indiqué appliquer des niveaux d'honoraires inférieurs pour les ONG et certains autres ANG que pour des bailleurs bilatéraux ou multilatéraux. La DGD n'était d'ailleurs pas tenue elle-même par cette limite.

Les ANG ont respecté le cadre réglementaire de la DGD en matière d'évaluation dans leur grande majorité. La très grande variété observée dans les évaluations (types, budgets, objets, objectifs, méthodologies, etc.) n'a généralement pas posé de problème en termes de respect du cadre réglementaire, dans le sens où celui-ci laissait globalement une grande liberté aux acteurs et en particulier aux ONG. Seulement une minorité de gestionnaires de la DGD (11% en moyenne sur tous les critères repris dans l'enquête) ont affirmé qu'il y avait des problèmes de respect des exigences « fréquents ou importants » pour les programmes dont ils avaient la charge. Les problèmes saillants étaient liés au respect des instructions formelles ou informelles sur les objectifs de l'évaluation (20%), le moment de l'évaluation (20%), l'objet de l'évaluation (13%) et les mesures pour favoriser l'indépendance de l'évaluation (13%). Une autre source de problème rapportée par de nombreux acteurs était la limite d'honoraires à 500€ (voir encadré 1). Plusieurs des points susmentionnés étaient en fait considérés comme peu clairs dans le cadre réglementaire ou ont fait l'objet d'instructions (personnelles) additionnelles au cadre réglementaire, comme décrit ci-dessus.

²¹ Ceci n'a pas posé de problème pour les universités et institutions scientifiques, dont les conventions prévoyaient une ligne budgétaire distincte pour les frais d'évaluation.

²² Le cadre réglementaire pour les « autres partenaires » a été adapté depuis lors, leur permettant depuis le 1^{er} janvier 2012 de budgétiser les frais d'évaluation conjointe ou externe comme des coûts opérationnels. Les ONG sont toutefois restées soumises à cette règle, vu la difficulté accrue de modifier un arrêté royal par rapport à des 'documents de référence'.

Encadré 2 – Distinction peu faite entre trois types d'objectifs

Trois principaux types de finalités peuvent être distingués pour l'évaluation, sans être exclusifs l'un de l'autre: (i) rendre compte ; (ii) aider à la décision ; et (iii) capitaliser sur les connaissances et partager les leçons apprises. Peu de distinction était généralement faite entre ces trois types d'objectifs.

Le premier objectif (« **rendre compte** »²³) vise à rapporter sur les résultats atteints par le programme, l'efficacité de sa mise en œuvre, etc. auprès des personnes en droit d'attendre ce type d'information. La reddition de compte peut ainsi s'entendre par exemple à la DGD, au Ministre, au Parlement, aux contribuables, aux pays partenaires, aux partenaires locaux, aux publics cibles destinataires, etc.

Le deuxième objectif (« **aider à la décision** ») vise à tirer les leçons d'une politique, d'un programme ou d'un projet, pour disposer d'une base factuelle solide pour aider (l'ANG, la DGD ou autre) à prendre des décisions liées à la suite à donner à cette action. Cela reste donc dans le cadre immédiat de la politique, du programme ou du projet évalué.

Le troisième objectif (« **capitaliser sur les connaissances et partager les leçons apprises** ») vise à tirer les leçons *transférables* d'une politique, d'un programme ou d'un projet, en vue de les partager avec un public plus large que celui du programme et de faire ainsi bénéficier d'autres politiques, programmes ou projets de l'expérience acquise.

La distinction claire entre ces trois types d'objectifs a rarement été rencontrée durant cette étude. Elle n'est pas faite dans le cadre réglementaire de la DGD (à l'exception partielle de celui pour les « autres partenaires ») et ne l'est pas non plus dans de nombreux rapports commandités par les ANG. Il ressort des entretiens avec la DGD et les ANG que le concept de cette distinction n'est pas répandu. Si la plupart des interlocuteurs ont affirmé que l'« apprentissage » était un objectif prioritaire pour l'évaluation, peu ont fait la distinction entre les deux types d'objectifs différents qui peuvent se cacher sous ce terme. Visaient-ils par exemple à tirer les leçons pour les partager et les diffuser avec un large public intéressé sur le sujet ? Ou plutôt à améliorer la formulation du programme suivant ? Il est important d'établir clairement quel objectif est visé en priorité, car cela a une série de conséquences sur la conception et la mise en œuvre de l'évaluation.²⁴ Vouloir poursuivre deux voire trois de ces objectifs en même temps entraîne le risque de disperser les ressources et les efforts ainsi que d'obtenir des effets contre-productifs.²⁵

²³ 'Accountability' en anglais - également appelée en français rendu de comptes, redabilité ou imputabilité.

²⁴ A titre d'exemple : une évaluation visant la reddition de compte, ou le partage des leçons, est typiquement centrée sur l'étude des résultats de l'action *passée* plus que sur l'étude du contexte et des enjeux *futurs* avec des recommandations d'action directement opérationnelles qui, elles, sont essentielles pour la prise de décision. A l'inverse, une étude focalisée sur la prise de décision ne nécessite pas d'examiner des résultats et leçons déjà connus en grande partie du management, alors que ceux-ci sont toutefois importants à étudier et exposer dans une optique de reddition de compte ou de partage de leçons.

²⁵ Des effets contre-productifs peuvent par exemple être observés en termes de moment et de durée du processus, de détermination du champ et des actions/pays à étudier et visiter, du choix des évaluateurs (p.ex. totalement indépendants ou connaissant justement très bien l'intervention à évaluer), etc.

3.1.2 Approche des ANG (Q1.2 et Q1.3)

Il y a des indications qui suggèrent que la capacité (institutionnelle et humaine) des ANG en évaluation durant la période 2008-2011 s'est améliorée dans un grand nombre de cas, entre autres grâce à la DGD pour la moitié des ONG. Ceci étant, près d'un tiers des ANG a connu une augmentation limitée.

- Une grande majorité (84%) des ANG considéraient que l'«évaluation» était bien (46%) voire très bien (38%) ancrée dans leur organisation.
- Deux tiers des ANG (63%) font part d'une augmentation tangible (52%) voire forte (10%) de leurs capacités en matière d'évaluation depuis 2007. Un tiers rapporte toutefois une augmentation limitée (32%). De nombreux ANG ont permis à un ou plusieurs membre(s) de leur personnel de suivre une formation traitant d'évaluation (voir l'encadré 3 ci-dessous).
- Lors de la constitution de l'échantillon pour cette méta-évaluation, 23% des ANG (15/64) ayant bénéficié de cofinancement de la DGD durant la période 2008-2011 n'ont pu fournir de rapport d'«évaluation» à proprement parler relatif aux activités cofinancées. Il s'agissait tous d'ONG, représentant ainsi un taux de 31% de celles-ci. A défaut, les rapports fournis dans le cadre de la « méthodologie d'évaluation » à définir par chaque ONG étaient par exemple ceux d'un exercice d'accompagnement, d'une étude de visibilité, d'une étude de ciblage, d'un exercice de synthèse, d'un effort de capitalisation, ou d'un exercice d'apprentissage stratégique (voir l'échantillon en Annexe 3).²⁶
- Plus de la moitié des ONG (54%) ont répondu à l'enquête que les exigences de la DGD étaient en partie (41%) à l'origine du renforcement des capacités de leur organisation en évaluation voire principalement (13%). Une telle contribution au renforcement des capacités, en tout ou en partie, est reconnue de manière moins importante par les 'autres partenaires' (27%) et encore moins par les universités et institutions scientifiques (17%).²⁷ Des gestionnaires indiquent toutefois que la DGD a contribué à ce que les universités développent leur approche méthodologique liée à la coopération au développement et fassent appel à de l'expertise hors du milieu académique (p.ex. experts en évaluation ou rompus à la coopération).
- Quelques acteurs avaient créé des mécanismes pour leur assurer un accompagnement plus structurel en évaluation, par un bureau d'études ou un « comité d'accompagnement des programmes » par exemple. Voir à ce propos l'encadré 6 dans la section 3.2.3
- Plus de la moitié des ANG (57%) ont indiqué dans l'enquête disposer à la fin 2012 de 1 à 4 personnes en interne maîtrisant les concepts et pratiques de l'évaluation

²⁶ Notons que l'inclusion de ces rapports dans l'échantillon n'a pas eu d'influence sur les constats sur la qualité. Il peut en effet être attendu que ces rapports répondent aux mêmes standards de qualité (description d'une méthodologie rigoureuse, argumentation solide des constats, conclusions valides, etc.). Des taux systématiquement plus élevés (même si légèrement) ont toutefois été observés dans la prise en compte des 'critères d'évaluation' (pertinence, efficacité, etc.) – voir section 3.2.1. La part relative minoritaire de ces autres types d'études explique cependant un faible impact des légères variations observées sur les constats généraux. De toute manière, l'examen de ces rapports est également amené à refléter la qualité de l'information générée pour tirer des leçons sur les résultats et la mise en œuvre des programmes cofinancés.

²⁷ Ce n'est pas tellement étonnant pour les partenaires universitaires et scientifiques dans la mesure où ils ont établi de longue date des pratiques d'évaluation. Il peut par contre étonner que les instructions limitées de la DGD pour les ONG eussent d'avantage contribué à renforcer les capacités en évaluation que les directives plus exigeantes pour les « autres partenaires », partenaires qui avaient également des pratiques et méthodes d'évaluation plus souvent peu définies (voir ci-dessous). Les informations dont nous disposons ne fournissent pas de clé de compréhension à ce propos.

(formations approfondies, conduite d'évaluations, etc.). Un tiers (32%) indiquait par ailleurs avoir plus de 5 personnes maîtrisant l'évaluation. Seuls 7 ANG (12%) ne disposaient pas de personnes expérimentées en évaluation. Aucun ANG n'a répondu à l'enquête n'avoir personne avec au moins des *connaissances de base* en évaluation. L'analyse statistique montre que le nombre de personnes avec des connaissances de base ou avancées en évaluation est, assez logiquement même si avec des exceptions, plus élevé quand les ANG ont plus de personnel. Il n'y a toutefois pas plus de quatre ANG qui ont mentionné avoir 10 personnes ou plus avec des compétences avancées en évaluation.

Encadré 3 – Formations suivies en évaluation

Des membres du personnel d'une grande majorité (80%) d'ANG ont pu bénéficier de formations traitant d'évaluation, en interne (43%) et/ou en externe (54%). Celles-ci étaient de qualité et de degré d'approfondissement variables.

Les **formations internes** étaient réalisées dans plusieurs cas de façon informelle ou via l'accompagnement d'experts en mission. Dans quelques cas il s'agissait de formations internationales à l'étranger (p.ex. par la Fédération Internationale de la Croix-Rouge) ou de formations organisées dans le cadre d'un processus d'accompagnement avec un bureau d'études (p.ex. Solidarité Mondiale). Les **formations externes** rapportées étaient souvent celles sur la gestion du cycle de projet organisées par les deux fédérations d'ONG, dont une partie abordait l'évaluation, ainsi qu'une évaluation spécifique en évaluation de la ngo-federatie. Il s'agissait également de formations données par exemple par MDF, INTRAC, ODI ou l'asbl STICS, ainsi que d'un cursus universitaire en évaluation de politiques publiques²⁸.

Cependant, la majorité des ANG n'avait pas défini de politique d'évaluation explicites et de guidance opérationnelle en la matière.

- Si de nombreux ANG avaient un certain niveau de *pratique* en évaluation, ceci ne s'est souvent pas accompagné d'une définition d'une *politique* et de *guidance* interne en la matière. Les trois quarts (74%) des ANG n'avaient ainsi pas élaboré de définition de leurs pratiques et méthodes internes en évaluation (44%) ou avaient certaines références à des pratiques et méthodes internes mais qui restaient assez générales (30%).²⁹ Ce constat est généralisé pour les 'autres partenaires' (91%), dont 73% ne disposaient pas d'une définition de leurs pratiques et méthodes (8 sur 11 ANG). Plusieurs ANG ont par ailleurs affirmé en entretien ou focus group ne pas savoir à quelle approche ou référence en particulier se raccrocher pour éventuellement définir une politique interne.
- 17 des 64 ANG avaient toutefois défini des directives internes relativement précises et formelles sur leurs pratiques et méthodes internes. Celles-ci ont même été considérées comme très précises et formelles pour 8 d'entre eux, dont 7 ONG, incluant entre autres trois grandes ONG internationales et une ONG ayant un partenariat avec un bureau d'études, mais également des ANG sans caractéristique distinctive particulière. Des pratiques intéressantes sont données à titre d'illustration dans l'encadré 4 ci-dessous.
- Notons enfin que si les deux tiers des ANG déclarent dans l'enquête utiliser leurs directives internes en pratique, un tiers a tout de même répondu à l'enquête qu'ils ne le font que peu ou très peu (16%) ou qu'ils 'ne savent pas' (16%).

²⁸ P.ex. le Certificat interuniversitaire en évaluation des politiques publiques en Communauté française de Belgique, organisé depuis 2011 par les FUCaM, l'UCL, l'ULB, l'ULG et la Société Wallonne de l'Évaluation et de la Prospective (SWEP).

²⁹ Ces chiffres proviennent de la synthèse des Fiches acteurs, qui sont le résultat de l'analyse des réponses aux enquêtes, des entretiens et focus groups, et des documents envoyés par les ANG en réponse aux demandes individuelles par email.

Encadré 4 – Politiques d'évaluation d'ANG intéressantes

Certaines ANG ont élaboré un guide d'évaluation interne, indiquant avec précision la politique d'évaluation de l'organisation et la méthodologie d'évaluation à appliquer en fonction des situations (p.ex. cadre logique, critères et indicateurs à évaluer, approches et outils d'évaluation, encadrement et contrôle de qualité, etc.). A titre d'exemple, Médecins Sans Frontières a créé une unité d'évaluation en 2005, qui a rédigé un manuel interne d'une trentaine de pages pour « initier, gérer et conduire des évaluations chez MSF ».³⁰ Un autre type d'exemple est le développement de pratiques avancées d'évaluation pour les divers programmes de coopération des universités flamandes par leur couple VLIR-UOS.

La part du budget des ANG consacrée à l'évaluation n'est pas aisée à établir, mais il y a des indications qu'elle se situerait en moyenne en-deçà du « 1% » de la période antérieure. De nombreux ANG nous ont indiqué garder un niveau de 1% du budget de leurs programmes à consacrer à l'évaluation comme repère interne dans leur organisation. Ceci s'inspire de la politique du « 1% » de la DGD entre 1997 et 2006, qui a contribué à insuffler la culture de l'apprentissage et de l'évaluation à de nombreux acteurs.³¹ Plusieurs de ces acteurs ont toutefois indiqué que le budget alloué à l'évaluation durant la période 2008-2011 était en fait souvent inférieur *in fine*, en raison d'arbitrages divers. Des gestionnaires de la DGD ont par ailleurs fait référence à une étude interne qui va dans le même sens, indiquant une moyenne d'environ 0,75%. La plupart des acteurs interrogés ne pouvaient toutefois pas avancer de chiffre précis, car rarement calculé en interne. Il faudrait également définir quels types de frais sont à considérer, comme des frais d'étude de faisabilité, d'évaluation ex ante, de définition de base line, de suivi (*monitoring*), de formation en évaluation, d'études de visibilité, d'évaluations multi-acteurs, de ressources internes, etc. ; ce type de difficulté avait déjà été soulevée dans l'étude d'évaluation du 1%.³² Certaines ONG avaient néanmoins défini une part plus élevée pour l'évaluation ; un taux de 2% a par exemple été rapporté dans plusieurs cas (pour des coûts de gestion de 7% du budget du programme).

Enfin, nous n'avons trouvé que **peu d'exemples d'évaluations couvrant des actions de plusieurs ANG**, telles des évaluations sur une thématique ou région particulière, comme cela s'était fait dans les années 1990 avec un comité comprenant la DGD (ex-AGCD) et des représentants d'ONG, d'après un interlocuteur. Quelques études commanditées les dernières années par le SES font exception.³³

³⁰ <http://Evaluation.msf.at/resources/reference-documents>

³¹ Comme décrit ci-dessus, cette politique a été abandonnée par la DGD en 2006 pour les ONG ainsi que pour les « autres partenaires », mais s'est maintenue pour les universités et certaines institutions scientifiques.

³² Iram, 2005. *Évaluation de l'utilisation du 1% du programme quinquennal des ONG consacré à l'évaluation, rapport de synthèse*. Service Public Fédéral Affaires Étrangères, Commerce extérieur et coopération au développement, Direction Générale de la Coopération au Développement (DGCD).

³³ P.ex. HIVA, ACE Europe et l'Institut de Politique et de Gestion du Développement (IOB), 2010. *Évaluation des partenariats des ONG orientés vers le renforcement des capacités*. Service de l'Évaluation spéciale de la Coopération au développement, Service Public Fédéral Affaires Étrangères, Commerce extérieur et Coopération au développement ; Channel Research, 2008. *Coopération belge au Rwanda, Évaluation 1994-2006*. Service de l'Évaluation spéciale de la Coopération au développement, Service Public Fédéral Affaires Étrangères, Commerce extérieur et Coopération au développement.

3.2 Qualité des évaluations – Q2

Les évaluations répondent-elles aux besoins et sont-elles de bonne qualité ? Quelle a été l'incidence du processus d'évaluation en matière de qualité de l'évaluation et d'apprentissage pour les parties prenantes ? Comment améliorer la qualité ?

Cette question vise essentiellement à déterminer le niveau de qualité des évaluations des programmes des ANG. La réponse à la question est structurée de la façon suivante : (i) la qualité des cahiers des charges ; (ii) la qualité des rapports d'évaluation ; et (iii) les facteurs explicatifs de la qualité des évaluations.

La réponse à cette question se base en grande partie sur l'analyse quantitative des appréciations des rapports d'évaluation et des pratiques des ANG. Pour rappel, ces appréciations sont elles-mêmes basées sur les indicateurs détaillés de la Matrice d'évaluation qui ont ensuite été croisés avec les réponses aux enquêtes auprès des ANG, des évaluateurs et des gestionnaires de la DGD. Les détails des chiffres fournis se trouvent dans les annexes.³⁴ Ils se basent sur l'analyse des rapports et ANG repris dans l'échantillon. Chacun des critères d'évaluation détaillés ci-dessous ont été appréciés pour eux-mêmes, sans qu'ils ne soient pénalisés par les appréciations données à d'autres critères (selon le principe de « non bis in idem »).³⁵

Q2 – Qualité des évaluations

Résumé de la réponse

La qualité des termes de références a été jugée complète et claire dans une large majorité de cas, traduisant l'effort mené par les ANG pour préciser leur demande. Ceci étant, par rapport aux concepts généralement admis en matière d'évaluation, les termes de référence présentaient une très grande hétérogénéité. Près d'un quart d'entre eux ne demandaient pas clairement d'examiner la pertinence du programme ni l'atteinte des résultats ; un tiers pas l'efficacité ; et la moitié pas l'impact et/ou pas la durabilité. D'autres critères et principes ont été très peu utilisés, comme ceux relatifs à la Déclaration de Paris ou aux Principes d'Istanbul.

Globalement, la qualité des évaluations pose problème. Dans la moitié des cas, elle a été jugée faible ou très faible, avec des différences en fonction du type d'ANG. Les faiblesses sont relatives à presque tous les critères de qualité. Les principaux facteurs expliquant les différences de qualité sont liés à la qualité de la demande, au budget trop faible, à l'absence de données de suivi, à la qualité des évaluateurs et à la qualité du processus (encadrement de l'évaluation).

³⁴ Annexe 5 (Matrice et résultats) ; Annexe 6 (Fiche Acteur et résultats) ; Annexe 7 (Fiche Rapport et résultats) ; Annexe 8 (Enquête ANG) ; Annexe 9 (Enquête évaluateurs) ; Annexe 10 (Enquête DGD)

³⁵ Selon ce principe, nul ne peut être jugé deux fois sur les mêmes faits.

3.2.1 Qualité de la demande et couverture de cette demande (Q2.1)

La qualité des termes de référence (TdR) est généralement bonne, même s'il y a des exceptions notoires pour plusieurs ONG.

- L'analyse montre que 80% des termes de référence (TdR) des rapports examinés peuvent être considérés comme complets et clairs, avec 26% de scores 'bon' et 54% de 'très bon' (!). Il ressort des entretiens et focus groups que la plupart des ANG réalisent en effet un réel effort au niveau de la formulation de la demande, parfois sur base de modèles standards développés en interne (constituant dans plusieurs cas l'essentiel de leur politique et pratique en évaluation).
- Les universités et institutions scientifiques se démarquent en particulier (100% de scores 'bon' ou 'très bon' !) ainsi que les autres partenaires (91%). Ces scores consistants pourraient s'expliquer entre autres par le fait que les TdR étaient typiquement revus par un comité de concertation ou de pilotage, incluant entre autres la DGD et parfois même (pour les syndicats) le SES.
- Les ONG avaient également des termes de référence de qualité dans une grande majorité des cas (75%). Un quart des TdR des ONG étaient toutefois faibles. Dans 9 cas (de différents types d'évaluations) il n'y avait en fait pas de TdR en bonne et due forme.

Les évaluations ont très souvent couvert la demande des TdR, en termes d'objectifs, de questions d'évaluation, de champ, etc.

- L'analyse montre que les évaluations répondent à la demande des TdR (objectifs, questions d'évaluation, etc.) dans 86% des cas et couvrent le champ formulé dans les TdR dans 95% des cas.
- Les évaluations ont largement été utilisées par les ANG en fonction des objectifs qui leur avaient initialement été assignés (voir également la Q6 sur l'utilisation par les ANG):
 - 65% avec objectif d'apprentissage → 58% utilisées comme tel
 - 41% avec objectif de reddition de compte → 45% utilisées comme tel
 - 83% avec objectif de prise de décision → 86% utilisées comme tel

Peu de consignes des TdR ont posé de problèmes spécifiques, d'après différentes parties, si ce n'est concernant les budgets de l'évaluation.

- **Budget** : les réponses à l'enquête auprès des évaluateurs indiquent que près de la moitié (43%) d'entre eux considérait que le budget était peu (37%) ou pas du tout réaliste (6%). L'analyse statistique sur base des 37 réponses ne fournit pas de typologie particulière. Néanmoins, il ressort des entretiens, des focus groups et des commentaires donnés en réponse aux enquêtes que le budget était souvent réduit par rapport à l'étendue de la demande (nombre de pays à visiter, de projets à examiner, etc.).³⁶ La méthodologie appliquée était ainsi souvent en deçà de ce qui aurait pu être réalisé pour un budget supérieur. De nombreux évaluateurs ont aussi indiqué qu'ils avaient *in fine* consacré beaucoup plus de jours à l'évaluation que ce que le budget prévoyait. La limite d'honoraires à 500 € de la DGD était par ailleurs souvent considérée comme irréaliste, voire contre-productive, et ainsi souvent contournée (voir Encadré 1 sous la Q1).

³⁶ Le budget des évaluations était très varié, allant de 2.000 € à plus de 200.000 €. 53% des évaluations de programmes d'ONG avaient un budget inférieur à 10.000€. Voir section 3.2.3

- Les consignes de **calendrier** ont posé relativement peu de problèmes (21% des ANG et 22% des évaluateurs considèrent que le calendrier était peu réaliste). L'analyse documentaire montre également qu'il y a nettement plus de cas où le calendrier a été globalement respecté (55%) que non respecté (14%).³⁷ Une difficulté mentionnée à plusieurs reprises est d'arriver à mobiliser des experts reconnus pour une demande à brève échéance et pour une charge de travail concentrée.
- Peu d'évaluateurs enquêtés considéraient que les consignes des TdR sur la **méthodologie** ou sur l'**équipe** ont posé problème (respectivement 16% et 6%).

L'analyse des TdR montre que des critères d'évaluation n'ont pas été utilisés de façon systématique. Les critères de durabilité et d'impact de l'OCDE-DAC l'ont par exemple été dans à peine la moitié de cas. Il était par ailleurs assez rarement demandé d'examiner des thèmes transversaux potentiellement importants ou intéressants.

- Les cinq critères d'évaluation de l'OCDE-DAC sont largement considérés comme les critères de référence pour évaluer l'aide au développement. S'il n'est nullement obligatoire ni nécessaire de les couvrir dans chaque évaluation, il est néanmoins intéressant d'examiner dans quelle mesure les ANG les ont considérés dans leurs TdR. Il ressort de l'examen des TdR que 73% demandent de couvrir la pertinence, 77% l'efficacité, 65% l'efficience, 53% l'impact et 58% la durabilité.³⁸ Vu inversement : près d'un quart des TdR ne demandent pas clairement d'examiner la pertinence du programme ni l'atteinte des résultats ; un tiers ne demandait pas l'efficience ; et la moitié ne demandait pas l'impact et/ou pas la durabilité.
- 88% des TdR ne font pas directement référence au processus sur l'efficacité de l'aide ayant mené à la Déclaration de Paris (2005), au Programme d'action d'Accra (2008) et au Partenariat de Busan (2011).³⁹ Il y avait également très peu de références aux huit principes d'Istanbul sur l'efficacité du développement des organisations de la société civile (2010).⁴⁰ Ceci n'a pas empêché certains TdR de référer à un ou plusieurs de ces principes, mais ce n'était en tout cas pas explicitement dans le cadre des processus internationaux susmentionnés.
- 73% des TdR ne définissaient pas de thèmes transversaux à examiner (genre, environnement, etc.).
- 94% des TdR ne demandent ni ne réfèrent à de l'innovation méthodologique, comme une approche d'évaluation innovante ou le développement d'outils méthodologiques transposables par la suite.

³⁷ Nous ne disposons pas d'information sur le respect du calendrier pour 32% des évaluations.

³⁸ Ces chiffres sont légèrement supérieurs (mais pas fondamentalement différents) s'il n'est tenu compte que des 51 explicitement appelées « évaluation » dans l'échantillon de 66 rapports (cf. Annexe 3 sur l'échantillon). Ceci donne 80% pour la pertinence, 86% pour l'efficacité, 73% pour l'efficience, 55% l'impact et 63% la durabilité. Les chiffres sur l'ensemble des 66 rapports montrent cependant bien le taux de prise en compte de ces critères par l'ensemble des ANG dans leur questionnement sur les programmes cofinancés, que ce soit par des études appelées 'évaluations' ou non.

³⁹ Les principes fondamentaux de la Déclaration de Paris sur l'efficacité de l'aide au développement étaient les cinq suivants : appropriation, alignement, harmonisation, gestion axée sur les résultats et responsabilité mutuelle.

⁴⁰ Les huit principes d'Istanbul ont trait aux thématiques suivantes : droits de l'Homme et justice sociale ; genre ; démocratie ; environnement ; transparence ; partenariat ; partage des connaissances ; et changement durable.

Notons par ailleurs que **tous les évaluateurs ayant répondu à l'enquête ont indiqué que la qualité des évaluations était une considération importante pour le commanditaire**, étant la priorité principale (41%) ou une priorité devant composer avec certaines contraintes (budgétaires, calendrier, intérêt du commanditaire, etc.).

3.2.2 Qualité des rapports d'évaluation (Q2.2)

Dans l'ensemble, l'appréciation générale de la qualité des rapports d'évaluation est faible dans pratiquement la moitié des cas. Si les évaluations sont jugées généralement bonnes pour les programmes des universités et institutions scientifiques, ceci n'est le cas que pour la moitié des programmes des ONG et des 'autres partenaires'. Le petit tableau ci-dessous résume ces appréciations.

Tableau 3 – Qualité des rapports d'évaluation, par type d'ANG

Appréciation générale de la qualité des rapports	1 - très faible	2 - faible	3 - bon	4 - excellent	Total
ONG	16%	37%	47%	-	(49) 100%
Universités et instit. scientifiques	-	17%	83%	-	(6) 100%
'Autres partenaires'	36%	9%	55%	-	(11) 100%
Total	18%	30%	52%	-	(66) 100%

Source : ADE

Les faiblesses en termes de qualité sont liées à pratiquement tous les critères examinés. Les problèmes les plus fréquents sont l'absence d'une revue critique de la conception du programme, des questions d'évaluation peu précises, l'absence de données de suivi fiables, le peu de précisions données sur le degré de fiabilité des constats et le manque de croisement des données. Autour de 40% des rapports d'ONG et d'autres partenaires sont par ailleurs faibles en termes de solidité de l'analyse et de validité et utilité des conclusions et recommandations. Un point positif récurrent (dans près de 80% de tous les cas) est néanmoins que les données sur lesquelles les évaluations se basent sont variées et équilibrées.

Ces différents points sont présentés dans le grand **tableau de synthèse ci-dessous**. Ce tableau montre l'appréciation synthétique des rapports par l'équipe d'évaluation sur base des informations provenant des différents outils d'évaluation (88 indicateurs de la matrice, enquêtes, entretiens, etc.). Notons que si un critère de qualité (ou plusieurs) a été considéré comme 'très bon' pour de nombreux rapports, aucun rapport n'a été apprécié comme de 'très bonne' qualité générale. L'appréciation générale s'est basée sur l'ensemble des critères, et pas sur un calcul mathématique. Il est en effet important de garder une capacité de jugement critique sur l'ensemble, tenant compte de l'importance relative des forces et faiblesses de chaque rapport (voir section 2.5).

**Tableau 4 – Appréciation de la qualité des 66 rapports d'évaluation
(synthèse sur base des Fiches Rapports)**

Degré d'appréciation : 1: très faible ; 2: faible ; 3: bon ; 4: excellent		1	2	3	4	TOT
Compréhension des Termes de Référence et couverture du champ de l'évaluation						
Réponse à la demande et couverture du champ	Le rapport reprend la demande formulée dans le cahier des charges (<i>objectifs, questions d'évaluation, etc.</i>) et y répond	-	14%	54%	32%	100%
	Le rapport présente une couverture complète du champ de l'évaluation (<i>champs temporel, géographique et thématique</i>)	-	5%	47%	48%	100%
	Le rapport témoigne d'une revue critique de la conception du programme (<i>questionnement du cadre logique, des liens de cause à effet attendus et des indicateurs choisis</i>)	28%	31%	25%	17%	100%
Qualité de l'approche méthodologique et du rapport d'évaluation						
Rigueur de la méthodologie	Le rapport indique que l'évaluation s'est déroulée suivant un processus phasé (par exemple: structuration, étude documentaire, visite de terrain, analyse et synthèse)	11%	18%	40%	31%	100%
	Le rapport présente des questions d'évaluation précises et adaptées au programme	22%	26%	31%	22%	100%
	Le rapport présente et justifie la méthodologie utilisée (<i>outils de collecte et d'analyse, sources d'information, limites de la méthodologie</i>)	6%	29%	53%	12%	100%
Fiabilité et représentativité des données collectées	Parmi les données collectées, certaines proviennent d'un système de suivi autorisant les comparaisons (<i>par exemple: existence d'indicateurs avec baseline et niveaux cibles</i>)	38%	45%	15%	2%	100%
	Les données collectées sont issues de sources d'information multiples et sont suffisamment larges pour tirer des constats généraux	2%	20%	50%	29%	100%
	Les données collectées sont équilibrées (<i>données primaires et secondaires, qualitatives et quantitatives</i>)	2%	20%	58%	21%	100%
	Le rapport précise les limites de fiabilité des données collectées	39%	33%	20%	8%	100%
Solidité de l'analyse	Le rapport témoigne de la triangulation des différentes données collectées	15%	27%	47%	11%	100%
	Les constats sont solides (<i>élaborés logiquement sur la base de données rendues explicites dans le rapport</i>)	9%	27%	59%	5%	100%
Validité des conclusions	Les conclusions sont valides (<i>étayées par des constats</i>) et répondent à la demande du cahier des charges	12%	27%	50%	11%	100%
Validité et caractère opérationnel des recommandations	Les recommandations dérivent des conclusions et sont opérationnelles (<i>précisant les actions à entreprendre, les acteurs responsables, le degré d'urgence, etc.</i>)	14%	23%	59%	5%	100%
Clarté de la présentation	Le rapport est organisé de manière claire et logique en différentes parties (<i>par exemple: méthodologie, constats, conclusions, recommandations</i>) et comporte un résumé	9%	26%	53%	12%	100%
	Le texte du rapport est aisé à lire de par sa clarté et sa concision	3%	27%	35%	35%	100%
	La présentation du rapport est attractive (<i>mise en page et longueur, tableaux, graphiques et autres illustrations</i>)	5%	33%	33%	29%	100%
Appréciation générale		18%	30%	52%	-	100%

Source : ADE

L'appréciation des différents critères présentés dans le tableau ci-dessus est détaillée ci-après.

Revue critique de la conception du programme

Plus de la moitié des évaluations n'ont pas examiné de façon critique le programme au niveau de sa conception.

- Le cadre réglementaire prévoyait que les programmes des ANG établissent un cadre logique pour leur programme. De nombreuses ONG ont également indiqué avoir internalisé la pratique des cadres logiques, notamment via la méthode de gestion du cycle de projet. Néanmoins, il ressort également des entretiens et enquêtes que les cadres logiques des programmes et/ou des projets ont fréquemment été remplis de façon formaliste, sans réflexion approfondie sur la logique inhérente à la chaîne de résultats escomptée ou sur la théorie du changement sous-jacente. Des projets s'inscrivaient par exemple dans une continuité historique, sans vraie remise en question de la pertinence de l'ensemble ou de certaines activités. Notons toutefois la pratique particulièrement intéressante développée dans le cadre de la coopération universitaire.⁴¹
- 58% des évaluations n'ont pas (28%) ou peu (31%) examiné la logique du programme/projet évalué et seulement 17% l'ont vraiment bien fait. Les ordres de grandeur sont les mêmes pour les différents types d'acteurs (ONG, universités et institutions scientifiques, autres partenaires). Certains l'ont fait au travers d'une reconstruction détaillée de la Logique d'intervention, d'autres au travers d'une question d'évaluation sur la pertinence de la conception du programme et des liens de cause à effet escomptés. Il ressort également des entretiens et focus groups que les acteurs ou évaluateurs n'ont souvent pas pensé à (faire) examiner la logique du programme.

Rigueur de l'approche méthodologique

L'approche méthodologique peut être considérée comme globalement rigoureuse dans environ deux tiers des cas. Le questionnement évaluatif était néanmoins souvent peu structuré et la méthodologie était décrite mais peu justifiée.

- 71% des évaluations se sont déroulées en différentes « phases »⁴², c'est-à-dire suivant une séquence logique avec des livrables spécifiques discutés avec le commanditaire et un comité de pilotage le cas échéant (par exemple : structuration → étude documentaire → visite de terrain → analyse et synthèse → dissémination). Pour les autres, il s'agissait souvent d'une succession d'activités sans livrable ou réunion, voire d'une mission d'évaluation d'une traite (p.ex. départ sur le terrain après rencontre du commanditaire pour en revenir directement avec le rapport final provisoire).
- 55% des évaluations ne présentaient pas de questionnement structuré avant de procéder à la collecte de données, tel que par une série de questions d'évaluation avec indicateurs spécifiques, tenant compte entre autres d'une revue des résultats

⁴¹ Des interlocuteurs ont mentionné que les académiques, peuvent avoir recours, outre à des formations de plusieurs jours au cadre logique, à des experts en cadre logique. Ceux-ci peuvent par exemple contribuer à l'atelier de préparation d'un programme, qui est typiquement prévu avec les partenaires locaux et au cours duquel le cadre logique est discuté. Un bond dans la qualité des programmes s'en est suivi, d'après ces interlocuteurs.

⁴² 40% des rapports reçoivent un score de 'bon' sur ce critère, 31% de 'très bon'.

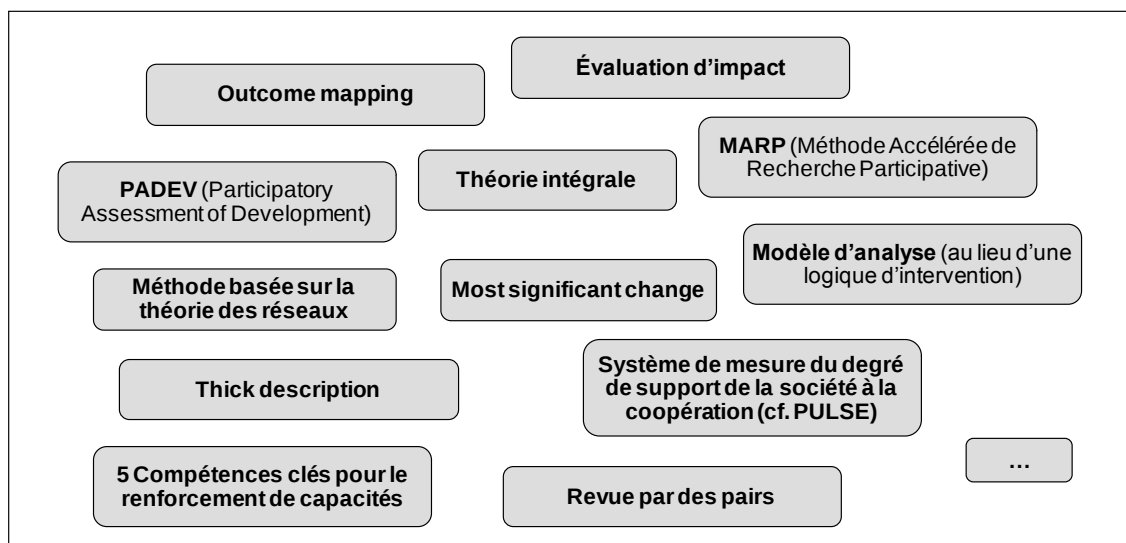
attendus. Dans 45% des cas il n'y avait pas de questions d'évaluation (21%) ou des questions pas clairement définies (24%).

- La méthodologie utilisée était souvent présentée, mais généralement avec peu de « justification » de l'approche générale et du choix des outils d'évaluation. Les limites n'étaient souvent pas (50%) ou peu (11%) décrites.
- Notons que 83% des ANG ont indiqué dans l'enquête qu'ils considéraient la méthodologie utilisée dans l'évaluation de leur programme comme (très) rigoureuse, un chiffre qui monte à 89% en demandant l'avis des évaluateurs. Le contraste avec notre analyse (globalement positive dans environ deux tiers des cas) n'est pas négligeable.

Si la DGD et les ANG n'étaient généralement pas demandeurs d'innovation méthodologique, des approches innovantes ont quand même été utilisées voire développées dans plusieurs évaluations.

- Le cadre réglementaire de la DGD n'incluait pas de clause particulière concernant l'innovation méthodologique ; il n'y avait pas non plus de mécanisme particulier permettant de promouvoir l'innovation. De nombreux interlocuteurs indiquent qu'au contraire, les gestionnaires de la DGD se montraient généralement réticents par rapport à des approches évaluatives non conventionnelles. Les ANG n'ont pas activement promu non plus l'innovation méthodologique, si on se base sur le fait que 94% des TdR ne la demandaient pas ou n'y référaient pas. Quelques initiatives ont toutefois été prises, comme la promotion du système de qualité EFQM par les deux fédérations d'ONG, même si ce système ne concerne pas en priorité l'évaluation externe.⁴³ Une ONG a également développé un processus d'accompagnement plus large avec un bureau d'études.
- Une dizaine des évaluations étudiées a néanmoins intégré des pratiques innovantes. 24% des 37 évaluateurs ayant répondu à l'enquête indiquent qu'ils ont utilisé des pratiques innovantes et 16% qu'ils en ont même développé dans le cadre de l'évaluation examinée. Il ressort des entretiens avec différentes parties que ces pratiques reflétaient l'intérêt de certains bureaux et évaluateurs pour l'innovation, mais également la nécessité d'approches spécifiques au vu d'un contexte particulier ou d'un budget limité, exigeant d'être créatif, même si parfois au détriment de la rigueur méthodologique. Des exemples de pratiques observés dans les évaluations étudiées sont présentés dans la figure ci-dessous et décrites dans l'Annexe 12. L'expérience avec ces pratiques innovantes fut tant positive que négative, en fonction des objectifs poursuivis et également des conditions de leur utilisation, dans le sens où plusieurs de ces pratiques nécessitent qu'elles fussent normalement intégrées au programme dès sa conception (comme l'évaluation d'impact ou l'*outcome mapping*).

⁴³ ACODEV et ngo-federatie promeuvent auprès de leurs membres depuis 2010-2011 le modèle de qualité EFQM, qui a été adapté aux particularités des ONG quand de besoin. Ce modèle créé par la Fondation européenne pour la gestion de la qualité (acronyme EFQM en anglais) est issu de la pratique du secteur privé. Il s'agit d'un cadre méthodologique pour permettre à ses membres d'évaluer leur niveau de qualité et de s'améliorer, par exemple en termes de leadership; de personnel, de stratégie, de partenariat et ressources, ou de processus, produits et services. Ceci vise à résulter en une meilleure performance et une plus grande satisfaction de toutes les parties prenantes. Les fédérations d'ONG rapportent que depuis juin 2011, près de la moitié des ONG avait adhéré à la démarche EFQM.

Figure 4 - Pratiques d'évaluation innovantes observées⁴⁴

Fiabilité et représentativité des données

Les données collectées dans les évaluations sont très souvent équilibrées et variées, mais leur fiabilité est limitée par l'absence fréquente d'un système de suivi objectif et de précisions claires quant à leurs limites.

- 13 évaluations (dont 12 d'ONG) sur 66 (20%) ont pu baser leur analyse sur un état des lieux antérieur au projet (« baseline study »).
- 73% des évaluations n'avaient pas (39%) ou peu (33%) à leur disposition de données provenant d'un système de suivi de projet comprenant des indicateurs et des niveaux cibles pour mesurer les effets.
- Les programmes triennaux des ANG dient néanmoins inclure une *baseline* et des indicateurs objectivement vérifiables (IOV), sur base duquel les ANG dient rapporter annuellement à la DGD sur l'ensemble du programme. Le peu d'utilisation de données de suivi semble donc liée à deux facteurs : (i) que les indicateurs du programme n'étaient pas suffisamment pertinents pour l'évaluation des résultats ou pas suffisamment spécifiques à l'objet de l'évaluation examinée (p.ex. pour l'évaluation d'une action dans un des pays soutenus) ; et (ii) que les ANG ne disposaient souvent pas d'un système performant de suivi et d'évaluation des résultats au niveau de leur programme ou des actions spécifiques.
- Point positif général : dans près de 80% de tous les cas les données sur lesquelles les évaluations se basaient peuvent être considérées comme variées et équilibrées. Ceci surtout en termes de mélange approprié de données de première main et de seconde main (82%), plus qu'en termes de données quantitatives et qualitatives (68%), ce qui rejoint en partie le point ci-dessus.
- Peu de rapports d'évaluation fournissent des précisions sur la validité des données, telles que sur les biais d'information (80% reçoivent un score faible ou très faible), les limites de fiabilité et d'utilité (67%), les règles de sélection de l'échantillon (45%) ou même les sources des données (29%).
- Notons que 92% des ANG ont indiqué dans l'enquête qu'ils considéraient les données utilisées comme « fiables et "représentatives" (sélection et taille de l'échantillon, existence d'une baseline, équilibre des données primaires/secondaires, variété des

⁴⁴ Voir Annexe 12 pour une description succincte de ces pratiques.

sources d'information, mention des limites, etc.) ». 78% des évaluateurs pensaient cela également. L'appréciation finale de ce critère sur base de nos outils d'évaluation, qui tient compte de ces réponses aux enquêtes, est tout de même moins positive.

Solidité de l'analyse et crédibilité des constats

L'analyse et les constats de deux tiers des évaluations sont considérés comme solides. De nombreuses évaluations présentaient des faiblesses en termes d'argumentation, d'analyse de causalité et de base de comparaison.

- L'analyse et les constats de deux tiers (64%) des évaluations sont considérés comme solides de façon générale.
- 79% des rapports d'évaluation font le lien dans les constats avec le contexte dans lequel le programme opère ;
- 68% des rapports d'évaluation présentent les constats de façon logique et justifiée (sans erreurs de raisonnement, d'inférences injustifiées, etc.) ; 62% montrent clairement le lien entre les données, l'analyse et les constats ; et 56% montrent que les données collectées ont été triangulées/croisées lors de la phase d'analyse.
- 56% des rapports ont des constats qui expliquent les liens de cause à effet entre les activités du programme et les effets observés (attendus ou inattendus).
- Une minorité des rapports ont des constats basés sur des comparaisons avant/après, bénéficiaires/non bénéficiaires, avec/sans, etc. (24%) ou sur une comparaison avec une baseline et/ou sur des données issues du système de suivi et d'évaluation (14%).
- Les sources d'information utilisées dans la justification des constats sont peu ou pas référencées dans 41% des rapports d'évaluation.

Validité et pertinence des conclusions

Les conclusions des rapports d'évaluation peuvent être considérées comme valides et pertinentes dans 61% des cas. De nombreux rapports n'ont pas de réelles conclusions ou ont des conclusions qui ne donnent pas de jugement d'ensemble clair et/ou d'argumentation solide.

- L'analyse montre que 61% des conclusions peuvent *in fine* être considérées comme « valides (étayées par des constats) et répondant à la demande du cahier des charges ». Ceci se base sur les observations suivantes, provenant de l'examen des rapports en tant que tels :
- Près d'un rapport sur quatre (24%) ne fournit pas de « conclusions » à proprement parler⁴⁵, se limitant par exemple à une énumération de constats par activité.⁴⁶ Ces cas ne sont pas spécifiques à certains types d'évaluation.
- Près de la moitié des rapports (44%) ont des conclusions ne donnant pas de jugement d'ensemble clair sur le niveau d'atteinte des objectifs du programme. Un

⁴⁵ Par 'conclusions' nous entendons des jugements d'ensemble clairs, basés sur les réponses aux questions d'évaluation, des constats et/ou de l'analyse. Les conclusions doivent avoir une valeur ajoutée par rapport à de simples faits ou constats.

⁴⁶ Ces rapports sont compris dans les statistiques présentées dans le reste du paragraphe sur les conclusions, dans lequel nous avons considéré les éléments de type conclusif ('conclusions', constats ou autres) dans notre appréciation des 'conclusions'.

même nombre (44%) n'a pas d'argumentation solide, montrant comment chaque conclusion s'appuie sur des constats.

- Néanmoins, l'enquête auprès des ANG a montré que 79% d'entre eux considéraient les conclusions de leur évaluation comme « (très) valides (base factuelle, évitant les biais de jugement, etc.) » et 88% comme « (très) utiles (couverture des points clés, intérêt stratégique et opérationnel, valeur ajoutée, etc.) ». Du côté des évaluateurs, 92% considéraient l'ensemble de leurs conclusions comme « (très) claires et étayées par des faits (mention des constats sous-jacents, évitement des biais de jugement, etc.) » et 89% comme « (très) utiles ».

Validité et utilité des recommandations

L'analyse des recommandations des rapports montre que celles-ci peuvent être considérées comme valides et opérationnelles dans 64% des cas.

- Comme pour les conclusions, l'appréciation finale des recommandations s'est faite sur base de la lecture des rapports et de l'avis des parties prenantes, en particulier des commanditaires.⁴⁷ Si la plupart des ANG se sont montrées satisfaites sur les recommandations (84% d'entre elles les considéraient comme (très) valides et utiles), notre analyse montre que 64% des recommandations peuvent *in fine* être considérées comme telles. Ceci se base sur les observations suivantes, provenant de l'examen des rapports en tant que tels :
- Près d'un rapport sur cinq (18%) ne fournit pas de recommandations.⁴⁸
- 61% des rapports ne montrent pas clairement comment les recommandations découlent des conclusions ou des constats.
- 62% des rapports précisent des options d'action et les acteurs responsables.
- 12% des rapports précisent l'urgence et/ou l'importance de chaque recommandation.

Clarté de la présentation

Les rapports avaient une présentation claire dans deux tiers des cas.

- 76% des rapports avaient une structure claire avec des sections dédiées aux constats, aux conclusions et aux recommandations. Ce chiffre comprend les nombreux rapports présentant chaque recommandation directement après la conclusion dont elle découlait.
- 68% des rapports étaient concis et faciles à lire, y compris 38% dont la lecture était très aisée.

⁴⁷ A la lecture initiale des rapports, l'équipe de la méta-évaluation a considéré que 50% des recommandations étaient globalement valides et opérationnelles. L'enquête auprès des ANG a montré que 84% d'entre elles considéraient les recommandations de leur évaluation comme « (très) valides et utiles (couverture des points clés, intérêt stratégique et opérationnel, valeur ajoutée, base factuelle, évitant les biais de jugement, etc.) ». Du côté des évaluateurs, 84% également considéraient l'ensemble de leurs recommandations comme « (très) valides et utiles » (3% des évaluateurs considéraient que leurs recommandations étaient d'utilité limitée ; les autres n'ont pas fourni de réponse ou ont répondu que la question n'était pas d'application, parce qu'il n'y avait pas de recommandations par exemple). En croisant ces différentes informations, ainsi que les entretiens éventuels, il ressort que 64% des recommandations peuvent *in fine* être considérées comme (très) valides et utiles.

⁴⁸ Ces rapports sont compris dans les statistiques présentées dans le reste du paragraphe sur les recommandations.

- 64% des rapports contenaient des annexes utiles, telles qu'une liste des sites ou pays visités, une liste des personnes rencontrées, une bibliographie ou des analyses spécifiques. Ce type d'informations manquait en tout (14%) ou en partie (21%) pour les autres.
- 59% des rapports avaient une mise en page, un graphisme et des visuels contribuant à leur clarté et les rendant attrayants à lire.
- 42% des rapports fournissaient un résumé complet, clair et concis, alors que 32 rapports sur 66 (58%) n'incluaient pas de résumé du tout.

Enfin, notons que, de façon récurrente, les ANG surestimaient la qualité des rapports d'évaluation, comme décrit dans l'encadré ci-dessous.

Encadré 5 – Surestimation de la qualité des évaluations par les ANG

Les ANG ont généralement surestimé la qualité du rapport des évaluations qu'elles ont commanditées, par rapport aux vues de la DGD et à l'analyse factuelle effectuée sur base des critères et indicateurs définis dans la matrice de qualité. Ceci peut s'expliquer principalement par le biais intrinsèque qui peut exister pour l'appréciation du travail sous leur responsabilité, ainsi que par le manque de connaissance des techniques d'évaluation, ou des attentes différentes en matière de qualité des évaluations. Une telle surestimation par les ANG a également été observée pour leurs capacités en évaluation (cf. section 3.1.2).

3.2.3 Facteurs explicatifs de la qualité (Q2.1, Q2.2 et Q2.3)

Les facteurs explicatifs de la qualité des évaluations sont liés à la demande, au budget, à la qualité des prestataires, et à la gestion du processus d'évaluation.

La demande

Comme signalé plus haut, **si la demande a été jugée claire et complète, les termes de référence témoignent dans un certain nombre de cas d'une connaissance limitée des concepts d'évaluation.** Une série de rapports ne pouvaient en fait pas même être considérés comme de l'« évaluation », ce qui explique certains scores faibles ou très faibles. Il y avait par exemple des rapports d'un exercice d'accompagnement, d'une étude de visibilité, d'une étude de ciblage, d'un exercice de synthèse, d'un effort de capitalisation, ou d'un exercice d'apprentissage stratégique (voir Annexe 3 – Échantillon).

Notons qu'il n'a pu être établi de lien statistique direct entre la qualité d'un rapport d'évaluation et la qualité des termes de référence de l'évaluation examinée. Ceci peut s'expliquer en partie par le fait que les scores de l'ensemble des rapports pour ce critère sur les TdR étaient relativement similaires, avec une grande majorité de 'bon' et de 'très bon', limitant ainsi le potentiel d'identification de corrélations statistiques.

Le budget

Le budget des évaluations était très varié, allant de 2.000 € à plus de 200.000 €, soit le centuple. Une majorité des budgets était toutefois faible : 58% des budgets des évaluations étaient inférieurs à 20,000 €, dont 39% du total inférieur à 10,000 € (voir le tableau ci-dessous).⁴⁹

⁴⁹ Les trois grands types de partenaires montraient chacun de grands écarts dans le budget des évaluations.

Tableau 5 – Budget des évaluations

Type	<10.000€	10.000€ - 20.000€	20.000€ - 40.000€	40.000€ - 80.000€	80.000€ - 200.000€	>200.000€	Sans réponse	Total
ONG	23	8	3	6	3	0	6	49
Univ&Scient.	1	1	2	0	1	1	0	6
Autres Part.	2	3	0	3	3	0	0	11
Total	26	12	5	9	7	1	6	66
ONG	47%	16%	6%	12%	6%	0%	12%	100%
Univ&Scient.	17%	17%	33%	0%	17%	17%	0%	100%
Autres Part.	18%	27%	0%	27%	27%	0%	0%	100%
Total	39%	18%	8%	14%	11%	2%	9%	100%

Source : ADE, sur base des 63 réponses à l'enquête auprès des ANG et d'information complémentaire

Il n'y a pas de corrélation statistique significative et forte entre le budget et la qualité des rapports. Ceci étant, **les évaluations jugées « très faibles » étaient dans 83% des cas réalisées avec des budgets inférieurs à 20.000 €**. Notons toutefois que des budgets faibles n'ont pas non plus empêché d'avoir des évaluations de qualité (17 rapports sur les 38 dont le budget était inférieur à 20.000 € ont été jugés de 'bonne' qualité).

La qualité des prestataires

Sans accès aux CV des experts en charge des évaluations, il nous était difficile d'apprécier directement la qualité de l'équipe d'évaluation.⁵⁰ L'analyse fait toutefois ressortir deux éléments qui influencent la qualité des rapports :

- **Les évaluations conduites par un bureau d'études (ou plusieurs) ont dans l'ensemble eu une proportion de rapports de bonne qualité clairement plus importante (63%) que les évaluations conduites par des évaluateurs contractés à titre individuel (36%),** comme indiqué dans le tableau ci-dessous. Les bureaux d'étude contractaient parfois eux-mêmes aussi des experts externes, même si dans les cas examinés ils ont souvent conduit les évaluations avec au moins en partie du personnel interne. Plusieurs facteurs pourraient ainsi expliquer les différences de qualité observées, notamment le fait que les bureaux disposent d'une expérience cumulée dépassant celle des experts contractés, de procédures de qualité plus systématiques et plus avancées (voir infra), ainsi que d'une plus grande capacité à solutionner des problèmes éventuels (p.ex. un remplacement d'expert ou des problèmes techniques et logistiques).

⁵⁰ L'équipe d'évaluation n'a pas eu accès aux offres des bureaux d'évaluation ou consultants indépendants répondant au cahier des charges des évaluations des programmes des ANG, étant donné le caractère sensible et confidentiel de la chose.

Tableau 6 – Qualité des rapports d'évaluation, par type d'évaluateurs

Appréciation générale de la qualité des rapports	1 - très faible	2 - faible	3 - bon	4 - excellent	Total
Bureaux d'études	8%	29%	63%	-	(38) 100%
Évaluateurs individuels	32%	32%	36%	-	(28) 100%
Total	18%	30%	52%	-	(66) 100%

Source : ADE

- **L'analyse statistique montre qu'il y a une corrélation forte et très significative⁵¹ entre la qualité générale des rapports et les garanties d'indépendance⁵², pour les ONG⁵³.** L'indicateur 'garanties d'indépendance' retenu dans notre analyse est basé sur une série d'éléments tels que l'exigence exprimée dans les TdR de l'indépendance des experts, l'exigence de fournir une déclaration d'absence de conflit d'intérêt, l'impartialité du choix de l'échantillon, l'accès des experts aux informations, l'absence de pression (information demandée aux évaluateurs au travers de l'enquête et lors d'entretiens), etc. Ces indices sont également le reflet du sérieux dans la recherche des experts et de la volonté de disposer d'un esprit critique, et donc d'une certaine garantie de la qualité des experts.

La gestion du processus d'évaluation

L'analyse statistique montre que **l'encadrement de l'évaluation par le commanditaire (p.ex. le suivi par un comité de pilotage) a globalement eu une influence limitée mais néanmoins positive sur la qualité générale des rapports.**⁵³

Un bon encadrement de l'évaluation n'a en fait été assuré par l'ANG que dans la moitié des évaluations examinées (52%). Les différences sont relativement marquées entre les ONG (43%), les universités et d'institutions scientifiques (83%) et les autres partenaires (73%). Dans les meilleurs des cas (12% ou 8/66, dont 5 'autres partenaires'), un comité de pilotage était mis en place, se réunissait à chacune des différentes phases du processus et faisait un contrôle de la qualité des tous les livrables, intermédiaires et finaux. Certains acteurs avaient même créé un mécanisme pour leur assurer un accompagnement plus structurel (voir l'encadré ci-dessous). Dans de nombreux autres cas toutefois, l'encadrement se limitait à une rencontre initiale avec le(s) évaluateur(s) et à une relecture du rapport final provisoire, avec des contacts informels plus ou moins fréquents avec les évaluateurs. Les différences observées entre les types d'ANG peuvent sans doute être expliquées dans une certaine mesure par les directives de la DGD, qui demandaient la constitution d'un comité de pilotage ou de concertation (incluant la DGD et dans certains cas le SES) pour les partenaires autres que les ONG, ainsi que par les pratiques propres à certains acteurs, comme les partenaires universitaires. Certains interlocuteurs ont également mentionné qu'ils

⁵¹ Corrélation forte (coefficient de corrélation de 0,562) et très significative (p-valeur de 0,01) ; il y a ainsi moins de 1% de chance d'obtenir les mêmes résultats sans corrélation.

⁵² Une telle corrélation est nettement moins marquée pour les universités et institutions scientifiques, ainsi que pour les autres partenaires, dans la mesure où pour ces acteurs-là les scores donnés aux garanties d'indépendance sont fort rapprochés.

⁵³ Corrélation significative (coefficient de corrélation de 0,072) mais relativement faible (p-valeur de 0.233).

considéraient un comité de pilotage trop lourd pour des évaluations à petit budget. Néanmoins, plusieurs ANG nous ont indiqué qu'ils utilisaient pour l'évaluation étudiée, dans une plus ou moins grande mesure, des comités créés dans un autre cadre, comme les comités de pilotage des programmes des 'autres partenaires' (requis par la DGD) ou un groupe d'experts externes contractés pour former un organe de réflexion interne.

Encadré 6 – Accompagnement structurel

Certains acteurs avaient créé des mécanismes pour leur assurer un accompagnement plus structurel. Une ONG avait par exemple fait appel à un bureau d'études pour un processus d'accompagnement incluant de l'évaluation et de la capitalisation. Une autre avait établi un « comité d'accompagnement des programmes » pour faire un suivi stratégique des programmes tout au long de leur durée. Il était composé de six experts indépendants contractés spécifiquement à cette fin. Ce comité a également été appelé à effectuer le suivi du processus d'évaluation.

Un peu plus de la moitié des évaluateurs ayant répondu à l'enquête indiquent que l'équipe d'évaluation (57%) avait un processus de qualité interne structuré, mais seulement 8% avaient un contrôleur qualité externe à l'équipe. 35% des équipes d'évaluation avaient un processus de qualité interne limité (24%) voire inexistant (11%).

Notons qu'un lien statistique direct n'a pu être établi entre la qualité d'un rapport d'évaluation et la définition par l'ANG d'une politique interne et de guidance opérationnelle en évaluation. Comme pour avec la qualité des TdR, ceci peut s'expliquer en partie par le fait que les scores de l'ensemble des rapports pour ce dernier critère étaient relativement similaires (généralement 'faibles' ou 'très faibles'), limitant ainsi le potentiel d'identification de corrélations statistiques. Il y a néanmoins des indications que **des acteurs aux pratiques d'évaluation bien ancrées et établies de longue date ont souvent eu des rapports de bonne qualité** (p.ex. universités et certaines ONG). Ceci s'explique notamment par le fait qu'ils ont souvent combiné des facteurs de qualité susmentionnés (budget, bureaux d'étude, encadrement, etc.).

3.3 Indépendance – Q3

Dans quelle mesure les évaluations sont-elles indépendantes?

L'indépendance de l'évaluation et la transparence du jugement donné dans les rapports d'évaluation peuvent avoir une influence sur la qualité et la crédibilité des évaluations. Cette question est donc étroitement liée aux questions Q2 et Q4. Elle vise à examiner l'indépendance sous un triple angle : (i) le caractère indépendant du lancement de l'évaluation et de la sélection de l'équipe d'évaluateurs ; (ii) le déroulement du processus d'évaluation sans pression ni entrave ; et finalement (iii) l'autonomie et l'absence de biais des évaluateurs dans la rédaction du rapport final. En effet, ces trois conditions doivent être réunies pour qu'une évaluation puisse être considérée comme indépendante.

Q3 - Indépendance

Résumé de la réponse

L'appréciation de l'indépendance a été réalisée à trois niveaux : l'organisation de l'évaluation et le choix des experts, la conduite des évaluations et l'examen des rapports.

Dans l'organisation des évaluations et le choix des experts deux éléments posent question. Premièrement, sur le plan des principes, le fait que les ANG étaient en charge des évaluations des interventions qu'ils avaient par ailleurs mises en œuvre. Si une telle auto-évaluation ne pose pas réellement problème et est même souvent justifiée dans une optique d'apprentissage et de prise de décision au niveau de l'intervention, elle soulève, en principe, des questions de conflit d'intérêt lorsqu'il s'agit de rendre compte ou de décider du renouvellement d'un financement. Ce conflit d'intérêt doit être géré soit par le recours à des évaluations indépendantes soit par un processus de contrôle et/ou d'accompagnement externes. Deuxièmement, en pratique les évaluations ont souvent été confiées à des 'évaluateurs externes', mais ceci ne signifie pas nécessairement qu'ils étaient indépendants. Dans le cadre de nos enquêtes, différents interlocuteurs ont soulevé la question du manque de recul et d'esprit critique d'experts qui venaient souvent du même milieu restreint des ANG belges.

La conduite des évaluations a également suscité des questions en termes d'indépendance, même si certains acteurs étaient d'avis qu'il y avait peu de difficultés de ce point de vue. Les ANG et les évaluateurs ont généralement affirmé que l'indépendance n'a pas réellement été mise en péril. Ce jugement général plutôt positif sur l'indépendance dans la conduite des évaluations n'est pas partagé par les gestionnaires de la DGD dont la majorité des répondants au questionnaire considéraient que certaines conditions d'indépendance n'étaient pas réunies. De plus, certains interlocuteurs font état de l'existence d'une autocensure ou de modification d'un rapport pour éviter la non continuation du financement.

L'examen des rapports d'évaluation ne suggère pas qu'il y ait eu un biais dans les rapports d'évaluation, mais que, au contraire, une large majorité de rapports (83%) présentent des jugements qui peuvent être considérés comme impartiaux. Ceci étant, le jugement ex-post sur la seule base des rapports du caractère impartial des jugements doit être pris avec précaution, un document ne suffisant pas à révéler s'il y a oui ou non eu des problèmes d'indépendance.

En conclusion, il ressort des différents constats que, même si certains éléments et certains acteurs ne pointent pas vers des problèmes d'indépendance, les doutes qui subsistent sont suffisants pour mettre en cause l'indépendance et porter atteinte à la crédibilité des évaluations menées.

3.3.1 Indépendance dans le lancement des évaluations (Q3.1)

Les différentes parties ont identifié l'indépendance comme une priorité, même si certaines étaient plus réservées sur ce point, notamment en affirmant qu'il y avait un arbitrage à faire entre indépendance et connaissance du sujet à traiter.

- Plus de 80% des gestionnaires de la **DGD** ayant répondu à l'enquête ont indiqué qu'ils estimaient, à titre personnel, que l'indépendance constituait la priorité (24%) ou qu'elle doit être recherchée en tenant compte des contraintes qui pèsent sur l'évaluation (59%). Certains ont précisé qu'au moins une évaluation externe indépendante était indispensable, afin d'obtenir un jugement réellement objectif. Trois gestionnaires (18%) ont répondu que l'indépendance n'était pas une priorité ou qu'elle était secondaire par rapport à d'autres priorités.
- Les chiffres sont assez similaires pour l'enquête auprès des **ANG** : près de 75% des répondants indiquaient que pour leur organisation l'indépendance était une priorité ou qu'elle était promue activement en tenant compte de contraintes ; 12% ont affirmé qu'elle y était secondaire par rapport à d'autres activités. Certains ont mis en avant les possibles limites de l'indépendance et notamment une moins bonne connaissance du contexte, de la technicité du sujet à traiter, ou encore de la spécificité de l'ANG concernée. D'autres par contre ont expliqué que les experts ne peuvent être impliqués dans les projets, au risque de perdre toute capacité d'avoir une vue externe objective ou encore que l'indépendance des évaluateurs était un critère de même importance que la connaissance du contexte.
- Près de la moitié des **évaluateurs** interrogés (43%) ont répondu que l'indépendance était la priorité, 22% qu'elle était recherchée de façon assez souple et 14% qu'elle était secondaire par rapport à d'autres priorités. Ici aussi, certains interlocuteurs ont fait référence à un arbitrage à faire entre indépendance et connaissance du sujet à traiter, en particulier lorsque l'objectif premier de l'évaluation était l'apprentissage.

Globalement, la mise en place des évaluations et la distribution des rôles posent question en termes d'indépendance pour principalement deux raisons : 1) le commanditaire était l'acteur qui avait également été en charge de l'organisation de l'évaluation du programme, dans la plupart des cas sans contrôle externe ; 2) certaines questions subsistent quant à l'indépendance de l'équipe d'évaluation.

- Comme expliqué dans la réponse à la Q1, le cadre réglementaire pour l'évaluation des programmes des acteurs non gouvernementaux attribuait la responsabilité de l'évaluation à ces acteurs, définissant eux-mêmes les termes de référence sur base desquels ils seraient évalués. Autrement dit, les ANG devaient évaluer les programmes qu'ils avaient mis en œuvre. Si cette approche peut être défendable dans un contexte où une évaluation poursuit d'abord un objectif d'apprentissage en vue d'améliorer le programme en question, elle peut être considérée comme une situation de **conflit d'intérêt** lorsqu'il s'agit d'informer la décision d'un renouvellement du programme ou encore lorsque l'évaluation vise avant tout à rendre compte.⁵⁴ « Aucune organisation n'est en effet encline à avouer ses propres

⁵⁴ On rappellera à ce sujet que les « *Principes du CAD pour l'évaluation de l'aide au développement* » précisent qu'on a le plus de chances d'avoir une évaluation impartiale et indépendante, en séparant la fonction d'évaluation de la fonction de gestion qui est en charge de planifier et gérer l'aide au développement (« *Impartiality and independence will best be achieved by separating the Evaluation function from the line management responsible for planning and managing development assistance.* ». OECD, DAC, 1991. *Principles for Evaluation of development assistance*, p. 6). De même, dans ses *Normes de qualité pour l'évaluation du développement*, le CAD stipule en termes d'indépendance des évaluateurs vis-à-vis des parties prenantes : « *Les évaluateurs sont indépendants de l'intervention d'aide au développement évaluée, y compris de ses fonctions de formulation politique, de fonctionnement et de*

fautes », comme mentionné dans un document du Parlement belge.⁵⁵ Il n'y avait par ailleurs ***pas ou très peu d'accompagnement externe*** qui aurait permis d'atténuer ce possible conflit d'intérêt. Comme montré dans la réponse à la Q1, le type d'accompagnement fourni par la DGD (par exemple la présence dans un comité de pilotage) variait en fonction du type d'ANG concerné, mais restait limité.

- Le cadre réglementaire fournissait certaines directives en termes de sélection des équipes d'évaluation, notamment en préconisant des évaluations externes. Comme expliqué dans la Q1, ***ces directives restaient toutefois relativement vagues*** et variaient en fonction des ANG concernés. Pour les ONG, par exemple, le cadre réglementaire référait implicitement à de l'évaluation externe, mais sans fournir davantage de précisions⁵⁶ ; pour les autres partenaires les exigences étaient plus détaillées.
- On notera encore à ce sujet qu'environ un tiers des gestionnaires de la DGD ayant répondu à l'enquête ont indiqué qu'il n'y avait pas (25%) d'exigences formelles (par exemple l'utilisation d'évaluateurs externes) de la DGD favorisant l'indépendance de l'évaluation ou des exigences peu claires et précises (6%), un quart indiquant qu'il y avait sur ce point des exigences tout à fait claires et précises et 38% les décrivant comme relativement claires et précises.
- De même, la majorité (65%) des Termes de référence ont obtenu l'appréciation « très faible » ou « faible » sur la ***présence de dispositions garantissant l'indépendance de l'évaluation***, plus précisément une sélection d'équipe garantissant une absence de conflit d'intérêt.
- Il n'en reste pas moins que sur base des réponses à l'enquête ANG on peut considérer que ***70% des évaluations ont été menées par des évaluateurs externes***, même si le sens exact de cette notion n'est pas toujours clair. Dans un peu plus de 20% de cas ils s'agissait de « pairs » (8%), d'« équipes mixtes » (8%) ou « d'évaluateurs internes » (5%), c'est-à-dire du personnel de l'ANG ou associé.
- De même, la large majorité des évaluateurs (92%) et des ANG (87%) ayant répondu aux enquêtes, ont affirmé qu'il n'y avait pas de lien fonctionnel ou hiérarchique entre les évaluateurs et le projet/programme à évaluer ou avec les acteurs de ce projet/programme. Ceci n'était toutefois pas formalisé : moins de 10% des répondants à ces deux enquêtes ont répondu que l'absence de lien était formalisée par la signature d'une déclaration d'absence de conflit d'intérêt.
- 94% des évaluateurs qui ont répondu à l'enquête considéraient qu'il n'y avait pas eu de soucis d'indépendance en matière de rédaction du cahier des charges et de sélection des offres et experts.

Mais, même dans les cas où les évaluateurs pouvaient être considérés comme externes (à l'ANG et ses programmes), l'indépendance d'esprit continuait à poser question pour certains, en termes de recul et de vision novatrice (*out of the box thinking*). Plusieurs personnes ***rencontrées*** considéraient qu'en fait ***commanditaires et évaluateurs faisaient partie d'un même cercle restreint***. Il est vrai que de nombreuses personnes ayant conduit les évaluations examinées étaient par exemple du personnel actif d'autres ANG, des anciens de l'ANG ou d'autres ANG, etc. D'après les témoignages

gestion, ainsi que des bénéficiaires prévus. » (OCDE, CAD, 2011. *Normes de qualité pour l'évaluation du développement*, p. 43).

⁵⁵ Chambre des Représentants de Belgique, 1997. *Suivi des problèmes de l'Administration Générale de la Coopération au Développement*, IV.42, p. 206. Même si cette citation ne concernait pas la coopération avec les ANG, mais la question de l'autonomie du service de contrôle et d'évaluation de l'AGCD de l'époque, elle est révélatrice des enjeux de l'indépendance.

⁵⁶ Ainsi cela peut signifier que l'évaluation doit être conduite par une équipe constituée entièrement d'externes, ou simplement que le processus soit suivi par une externe ; ou encore s'agit-il d'experts qui n'ont pas été impliqué dans ce programme particulier, ou dans aucun programme de l'ANG, etc.

des ANG, c'était également souvent un même petit nombre de bureaux d'étude belges qui étaient sollicités pour faire offre.

3.3.2 Indépendance dans la conduite des évaluations (Q3.2)

Les différents acteurs n'ont pas les mêmes jugements globaux en termes de l'indépendance du processus d'évaluation : les représentants de la DGD émettent des réserves quant à l'indépendance du processus, tandis que les évaluateurs et surtout les ANG font moins état de difficultés de ce point de vue. Les exemples montrent toutefois que le processus d'évaluation a dans certains cas, posé problème en termes d'indépendance.

- Environ un tiers (5 personnes) des gestionnaires de la **DGD** ayant répondu à l'enquête ont affirmé que toutes les conditions étaient réunies pour assurer l'indépendance de toutes les évaluations, mais 67% considéraient que des conditions n'étaient pas réunies pour certaines évaluations. Parmi les problèmes évoqués, on notera le fait que des évaluations avaient été menées par des pairs (des ONG entre elles), voire par d'anciens collègues. De même, plusieurs interlocuteurs ont fait état d'une autocensure et que le comité de pilotage avait eu un rôle important à jouer pour assurer l'indépendance, ou encore que les ANG avaient parfois tendance à chercher des évaluateurs « *qui ne font pas de vagues* ».
- Parmi les **évaluateurs**, aucun des répondants n'a dit qu'il y avait eu de fortes contraintes ayant affecté l'indépendance, 77% affirmant qu'il y avait une « indépendance totale » et 20% disant qu'il y avait indépendance en dépit de certaines contraintes. Les évaluateurs ont toutefois aussi fourni des exemples de situations qui posaient clairement problème en termes d'indépendance. Dans un cas précis il a été spécifié qu'un rapport a été adapté car il dit être envoyé à la DGD et que des éléments du rapport auraient pu entraîner la non continuation du financement du programme. Dans un autre cas, un évaluateur a signalé que la synthèse envoyée à la DGD ne contenait pas certains aspects négatifs mentionnés dans le rapport.
- De manière assez similaire, 82% des **ANG** ayant répondu à l'enquête ont dit que l'équipe d'évaluation a eu une « indépendance totale » pour conduire l'évaluation et écrire son rapport ; 15% estiment qu'il y a eu certaines contraintes de ce point de vue. Certains fournissent des exemples d'évaluations participatives qui ont mené à des blocages, mais la plupart ne semblaient pas considérer qu'il y ait eu des problèmes relatifs à l'indépendance, même dans les évaluations participatives.

3.3.3 Indépendance des rapports d'évaluation (Q3.3)

Il ressort de l'analyse des rapports que, pour la plupart d'entre eux, les constats, conclusions et recommandations étaient objectifs et non-biaisés. En effet, pour 83% des rapports analysés, l'équipe a considéré que les jugements présentés pouvaient être considérés comme impartiaux. Ce chiffre est similaire pour les ONG (84%), plus bas (73%) pour la catégorie « autres partenaires », mais il est de 100% pour les universités et institutions scientifiques. Dans les cas où l'impartialité était mise en doute, c'était par exemple dû au fait que certains jugements ne fournissaient pas de justification convaincante sur base de faits établis dans le cadre du programme évalué.

Ces chiffres doivent toutefois être traités avec précaution, car il est particulièrement difficile de juger de l'indépendance d'une évaluation en se basant uniquement sur un rapport. En effet, les biais sont difficilement identifiables. Ils peuvent également être dus à d'autres causes que le manque d'indépendance et l'absence de biais n'est pas une garantie d'indépendance. Néanmoins, un rapport peut révéler des problèmes d'indépendance, par exemple en les mentionnant explicitement ou parce qu'il contient certains biais qui suggèrent que l'équipe a été mise sous pression

pour modifier certaines parties du rapport (sans que ces modifications aient été effectuées pour corriger des erreurs factuelles, des fautes de raisonnement, l'oubli de certaines données, etc.) Cela peut être le cas lorsque des conclusions ne semblent pas s'appuyer sur des constats, ou encore lorsque des recommandations ne semblent pas cohérentes avec ce qui est stipulé dans les conclusions. Cela étant dit, il est souvent difficile de déceler ce type de biais, qui sont généralement subtils. Ces biais peuvent par ailleurs être dus à d'autres causes, comme un manque de rigueur de l'évaluation. Enfin, ce n'est pas parce que des rapports ne révèlent pas de tels biais qu'on peut conclure qu'il n'y a pas eu de problèmes d'indépendance : un rapport peut très bien ne pas mentionner une information qui modifierait l'appréciation sur un point particulier, sans que ceci soit apparent dans le rapport. L'absence de biais apparent n'est ainsi qu'une indication parmi d'autres ; elle ne suffit pas pour conclure à l'indépendance d'une évaluation.

3.4 Crédibilité – Q4

Quelle est la crédibilité des évaluations?

La crédibilité des évaluations dépend d'abord de leur qualité : plus une évaluation est de bonne qualité, plus elle est crédible. Le caractère indépendant des évaluations peut également contribuer à la crédibilité des évaluations auprès des audiences extérieures. Enfin, par leur capacité à contribuer à un climat de confiance, l'implication des parties et une communication transparente au sujet de l'évaluation peuvent favoriser la crédibilité. Cette question est donc, entre autres, étroitement liée aux deux questions précédentes. Elle est traitée en examinant la crédibilité : (i) des évaluateurs, (ii) de la méthodologie, (iii) du processus, (iv) des livrables et (v) de la communication des résultats finaux.

Q4 - Crédibilité

Résumé de la réponse

La crédibilité des évaluations est souvent mise en doute, comme l'attestent notamment les répondants aux questionnaires DGD et les analyses effectuées par l'équipe d'évaluation.

Plusieurs éléments ont pu affecter cette crédibilité :

- La crédibilité des parties a pu être mise à mal par les réserves en termes d'indépendance des commanditaires et des évaluateurs mentionnées dans la réponse à la Q3. L'indépendance peut toutefois avoir des effets opposés en termes de crédibilité en fonction des parties prenantes : pour le citoyen contribuable l'indépendance peut être un gage de crédibilité ; tandis que pour le gestionnaire du programme elle peut précisément diminuer la crédibilité. Le professionnalisme des évaluateurs ne semble pas avoir été mis en cause, ni d'ailleurs la fiabilité ou l'indépendance des personnes ressources consultées.
- Les faiblesses méthodologiques des rapports d'évaluation, le manque de recours à un contrôle de qualité externe, le manque de moyens financiers alloués à l'exercice d'évaluation et le manque de communication sur les résultats sont des facteurs qui ont pu avoir un impact négatif en termes de crédibilité.
- La consultation des partenaires, généralement répandue, a par contre pu avoir des effets favorables en termes de crédibilité.

De manière générale, **la crédibilité des évaluations a souvent été mise en doute** :

- En effet, si 65% des répondants DGD ont indiqué que la majorité des évaluations étaient crédibles, 30% ont affirmé soit qu'elles étaient majoritairement peu crédibles ou peu crédibles dans un ou plusieurs cas. Ce chiffre n'est pas marginal. Une évaluation qui n'est pas jugée crédible par une personne perd une part importante de son utilité pour celle-ci.
- L'analyse des rapports effectuée par l'équipe d'évaluation va dans le même sens. En effet, l'appréciation globale est que la crédibilité était bonne dans 58% des cas mais faible ou très faible dans 42% des cas.

La crédibilité d'une évaluation dépend de différents facteurs, qui sont examinés un à un ci-après.

Les parties sont-elles crédibles ?

Pour qu'une évaluation soit crédible, il faut que les parties à l'évaluation le soient également. Ceci concerne, principalement, trois types : le commanditaire, l'évaluateur et les parties consultées pour collecter de l'information. Celles-ci doivent être crédibles sous l'angle de l'indépendance. Elles doivent également l'être en termes de « qualité », c'est-à-dire de perception du « professionnalisme » du commanditaire et de l'évaluateur. Il y a enfin la question de la fiabilité, par exemple des sources ou des personnes fournissant des informations.

La réponse à la Q3 a montré que l'indépendance des commanditaires et des évaluateurs posait question, ce qui a pu affecter la crédibilité des évaluations. Il est toutefois important de souligner que l'indépendance peut avoir une incidence différente en termes de crédibilité en fonction de la partie prenante concernée. En effet, pour des parties prenantes plus ou moins externes au programme (par exemple le financeur ou le citoyen contribuable) l'indépendance peut être un gage de crédibilité. Par contre, on comprendra aussi que l'indépendance peut avoir des effets inverses en termes de crédibilité pour les parties qui sont directement liées au programme. Ainsi, et cela a aussi été évoqué par plusieurs de nos interlocuteurs dans le cadre de cette méta-évaluation, les parties prenantes proches du programme peuvent considérer que la crédibilité d'un évaluateur entièrement externe est moindre, considérant qu'il ne peut avoir une connaissance suffisante du contexte ou du programme ou encore qu'il n'effectuera pas l'évaluation dans un esprit de partenariat, mais uniquement en vue d'émettre un « jugement ».

Nous ne disposons pas d'information concernant la fiabilité ou le « professionnalisme » des commanditaires. Par contre, comme expliqué dans l'encadré ci-après, de nombreux ANG ont eu des difficultés à trouver de bons évaluateurs ou des évaluateurs tout court. Cela étant dit et même si la qualité des évaluateurs était jugée parfois inégale, la grande majorité des ANG ayant répondu à l'enquête (89%) considéraient que les évaluateurs avaient joué un rôle déterminant dans la qualité des évaluations.

Enfin, peu d'information est disponible sur la fiabilité ou l'indépendance des personnes ressource consultées. Cela étant dit, l'information collectée ne fait pas mention de problèmes majeurs de ce point de vue.

Encadré 7 – Identification des évaluateurs

De nombreux ANG ont fait part de leur difficulté à trouver de (bons) évaluateurs.

L'identification d'évaluateurs externes (d'individus ou de bureaux d'étude) spécialisés dans le domaine d'évaluation requis n'a souvent pas été aisée, en particulier pour les acteurs qui n'avaient qu'une expérience récente de l'évaluation externe. Ils ne disposaient pas de mécanismes ou d'outils particuliers à leur disposition, tel qu'un réseau d'experts ou une base de données (comme il en existe une à usage interne à la DGD concernant des experts sectoriels, même si celle-ci n'a pas été tenue à jour). Ce problème a été soulevé dans certains domaines de coopération spécifiques, comme la coopération syndicale.

La qualité du travail des évaluateurs était inégale, ce qui est apparu clairement dans des évaluations de synthèse sur base de rapports pays. Toutefois, seuls 8% des ANG ayant répondu à l'enquête ont indiqué que les évaluateurs avaient eu un rôle négatif déterminant dans la qualité de l'évaluation, 33% indiquant un « rôle positif » et 56% un « rôle très positif ».

Les évaluations reposent-elles sur des méthodologies d'évaluation crédibles ?

La réponse à la Q2 se montre positive en termes de couverture du champ par les évaluations.

Elle est par contre plus critique sur la rigueur et transparence de l'approche méthodologique utilisée, pointant vers des faiblesses en termes de phasage des évaluations, utilisation d'un questionnaire structuré, justification de l'approche suivie et des outils utilisés et vers un manque de référence aux limites de l'évaluation. Pour 41% des rapports, l'équipe d'évaluation a d'ailleurs estimé « faible » ou « très faible » la solidité de l'analyse et la crédibilité des constats (voir Q2). Les ANG et les évaluateurs par contre qualifiaient généralement les méthodologies utilisées de rigoureuses.

Le processus de l'évaluation a-t-il contribué à la crédibilité de l'évaluation ?

Le bilan en termes de **consultation des différents types d'acteurs** du projet/programme est assez bon. En effet 83% des répondants au questionnaire ANG ont indiqué que les parties prenantes ont été impliquées fortement tout au long du processus (16%) ou aux étapes clés (67%). Les rapports donnent une image moins positive :

- Environ 65% des rapports mentionnent que le processus d'évaluation visait à sensibiliser et à impliquer les parties dans l'évaluation, mais dans un peu plus de 1/3 des cas rien n'était dit à ce sujet ;
- La moitié des rapports d'évaluation indiquent que le processus d'évaluation a permis aux acteurs de l'intervention d'identifier des bonnes pratiques et des leçons apprises, mais dans un peu plus de 40% des cas ceci ne ressort pas.

Par contre, l'enquête auprès des évaluateurs montre que peu d'évaluations ont mis en place un **processus de qualité** très structuré, que ce soit au sein de l'équipe ou dans le chef du commanditaire (voir le tableau ci-après).

Tableau 7 – Processus de qualité structuré

	Équipe d'évaluation	Commanditaire
Processus de qualité structuré et contrôle de qualité externe	8%	3%
Processus de qualité structuré	49%	41%
Processus de qualité limité	24%	19%
Absence de processus de qualité	11%	16%

Source : ADE, sur base des 37 réponses à l'enquête auprès des évaluateurs

En effet, environ la moitié des répondants fait état d'un processus de qualité structuré, tant au sein de l'équipe que dans le chef du commanditaire. Le recours à un contrôleur de qualité externe à l'équipe n'est mentionné que par trois répondants (8%). Cette pratique est pourtant relativement courante et une exigence systématique pour certains commanditaires expérimentés. Seul un répondant fait état du recours par le commanditaire à un contrôleur qualité externe.

En termes de **moyens** consacrés à l'évaluation, on retiendra de la réponse à la Q2 que :

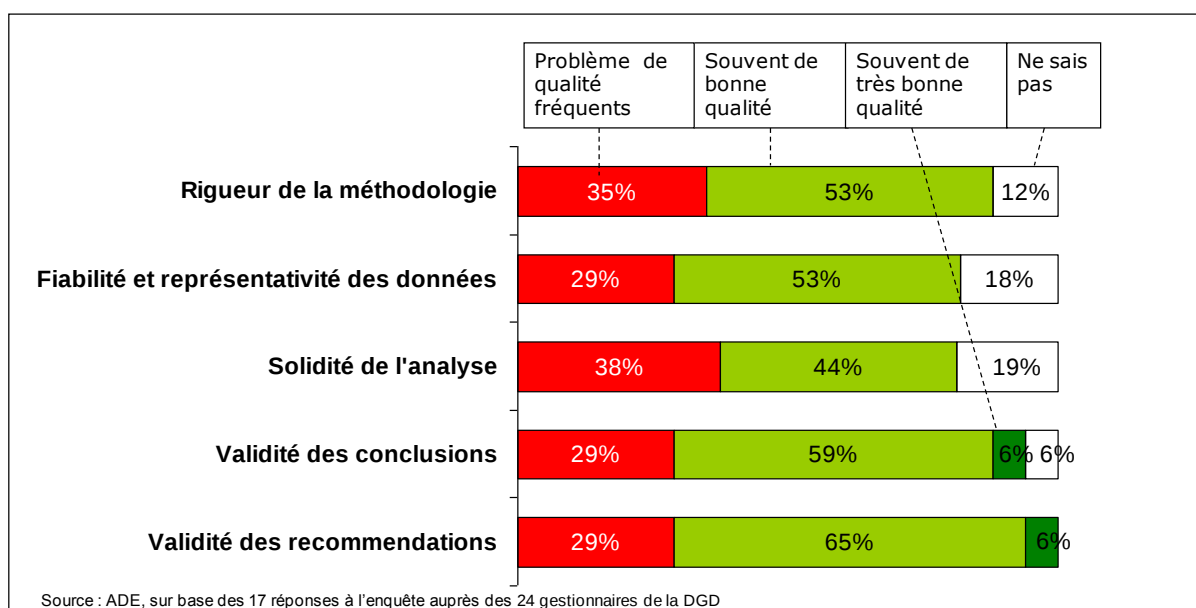
- De manière générale, plusieurs éléments indiquent qu'il y avait une discordance entre les budgets et les exigences formulées dans les TdR des évaluations; certains budgets étaient insuffisant pour une évaluation de qualité ce qui a pu affecter la crédibilité.
- Les calendriers n'ont généralement pas posé de problème.

Le livrable (rapport final) est-il crédible ?

Comme noté plus haut et en réponse à la Q2, les rapports posent problème sur plusieurs aspects touchant à la qualité. Ceci a affecté la crédibilité des évaluations, comme confirmé par le questionnaire DGD qui montre qu'environ un tiers des répondants ont estimé que ces divers critères ont posé problème. Par ailleurs, l'analyse a montré qu'il y avait une corrélation très forte⁵⁷ entre leur appréciation en termes de qualité sur les différents critères et leur appréciation en crédibilité telle que mentionnée plus haut. Ceci pourrait indiquer que ce sont les faiblesses en qualité qui expliquent leur appréciation en termes de crédibilité.

⁵⁷ Coefficient de corrélation de 0,779.

Figure 5 - Appréciation des répondants de la DGD sur différents critères de qualité



Par ailleurs (voir également la Q2), deux tiers des rapports étaient considérés par les gestionnaires de la DGD comme clairs en termes de présentation. Les chiffres sont similaires en matière de présence d'annexes utiles pouvant contribuer à la transparence et donc à la crédibilité des rapports.

La stratégie de communication des résultats de l'évaluation contribue-t-elle à sa crédibilité ?

Peu d'éléments permettent de lier stratégie de communication des résultats et crédibilité des évaluations. On notera toutefois qu'environ la moitié (55%) des rapports examinés par l'équipe indiquent qu'une restitution des résultats de l'évaluation après remise du rapport final a eu lieu (ou est prévue) au Nord et/ou au Sud. De même pour environ la moitié des évaluations (47%), l'ANG a produit un document de positionnement sur les recommandations de l'évaluation dans la foulée du rapport d'évaluation. Dans un tiers des cas, l'ANG a décrit un plan d'action pour mettre en œuvre les recommandations qu'elle juge utiles.

Par contre la pratique consistant à rendre publics les rapports d'évaluation n'est pas très répandue. En effet, l'enquête auprès des ANG a montré que le rapport a été régulièrement diffusé auprès des parties prenantes très directement concernées (89% des répondants disent que le rapport a été envoyé au gestionnaire de la DGD et 76% aux acteurs du projet/programme). La diffusion à un public plus large est bien moins souvent mentionnée (25% pour la diffusion auprès de membres de l'ANG, les autres ANG, des destinataires d'une lettre d'information). La publication sur un site web ouvert est évoquée par 10% des répondants. Un peu moins d'un tiers des répondants (27%) parle d'une diffusion plus large d'une partie du rapport (par exemple le résumé). Ces chiffres sont similaires pour les différents types d'ANG (voir aussi la section 3.6.1).

3.5 Contenu des évaluations – Q5

Y a-t-il des constats, des conclusions ou des recommandations qui reviennent fréquemment dans les rapports des acteurs non gouvernementaux ? Quels sont les principaux points positifs ou négatifs de l'ensemble des rapports d'évaluation ?

Cette question porte sur le contenu des évaluations. Elle vise à présenter les principales conclusions et recommandations qui ressortent des différents rapports d'évaluation en vue d'analyser les principales forces et faiblesses du travail des ANG. L'analyse présentée dans cette question se fonde sur la grille de contenu récurrent qui dresse un aperçu des conclusions et recommandations récurrentes. Elle suit étroitement la structure adoptée dans cette grille (voir Annexe 11). Elle est ainsi articulée autour des cinq critères d'évaluation du CAD de l'OCDE (la pertinence, l'efficacité, l'impact, la durabilité et l'efficience), ainsi qu'autour de la coordination et des thèmes transversaux.

Q5 – Contenu des évaluations

Résumé de la réponse

Il est difficile d'identifier des conclusions et recommandations récurrentes dans les rapports des ANG. En effet, ces rapports n'ont pas été conçus pour consolider l'information qu'ils contiennent : ils portent sur des interventions de nature très diverses et fournissent des conclusions et recommandations qui sont très spécifiques à ces interventions et par là même difficilement généralisables.

Il est toutefois possible de déceler certaines tendances en restant au niveau des critères d'évaluation et des thèmes d'analyse classiques. De manière générale, les conclusions et recommandations des rapports concernent principalement les problématiques de pertinence et d'atteinte de résultats. Certaines thématiques comme l'impact ou les thèmes transversaux sont rarement traités.

Parmi les points positifs soulignés dans les conclusions de nombreux rapports, on retiendra le fait que les programmes répondaient à des besoins clairement identifiés et étaient cohérents avec les stratégies nationales des pays bénéficiaires. De même les rapports affirment souvent que les résultats attendus ont été atteints ou l'ont été pour certains d'entre eux, même si cette affirmation concerne en général des réalisations, voire des activités. Ils sont par contre plus critiques sur la conception formelle des stratégies qui a posé problème dans de nombreux cas, ou encore sur les systèmes mis en place pour mesurer les résultats. Même si la durabilité n'est pas toujours traitée, les rapports sont plutôt négatifs sur ce point, soulignant par exemple l'absence d'une stratégie de sortie comme risque majeur.

Les recommandations récurrentes portent sur la nécessité de mieux inscrire les projets dans leur contexte et d'améliorer la qualité du cadre logique. De nombreux rapports soulignent aussi l'importance d'améliorer les systèmes de suivi et évaluation. Enfin, en termes de durabilité, les rapports soulignent surtout l'importance d'assurer un financement continu.

3.5.1 Remarques préliminaires

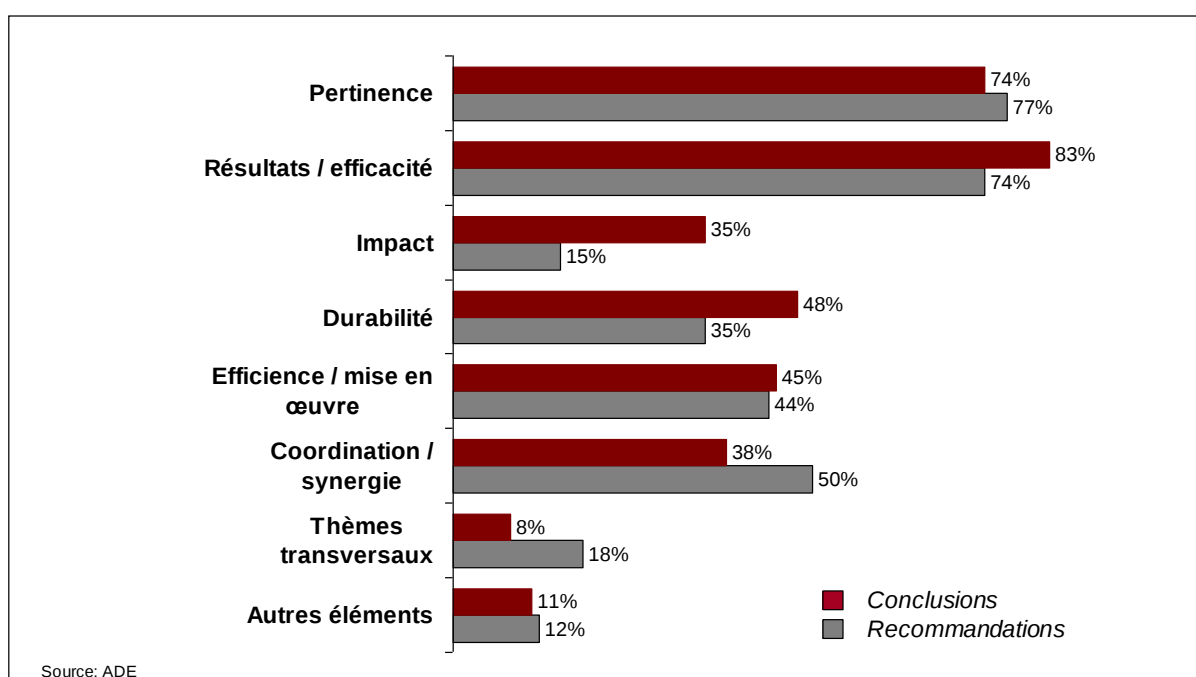
A la lecture des rapports d'évaluation, plusieurs constats généraux s'imposent quant à leurs conclusions et recommandations :

- Les évaluations concernent des programmes de secteurs très divers ;
- Elles évaluent ces programmes sans suivre un canevas commun qui permettrait, par exemple, d'identifier des questions ou thèmes qui devraient être traités par toutes les évaluations ;
- Les conclusions et recommandations restent généralement très proches des spécificités des programmes et fournissent peu d'éléments qui sont généralisables (i.e. qui pourraient par exemple s'appliquer à d'autres programmes), ce qui ne met pas en cause l'utilité de ces conclusions et recommandations.

Dans ces circonstances, si l'examen des rapports permet d'identifier des forces et faiblesses récurrentes du travail des ANG, celles-ci consistent en des jugements généraux relatifs aux critères d'évaluation du CAD de l'OCDE, la coordination et les thèmes transversaux. Les conclusions et recommandations sont trop disparates pour permettre d'avoir une vue plus détaillée et fiable de ces principales forces et faiblesses.

On notera par ailleurs que les rapports d'évaluation offrent surtout des conclusions sur la pertinence des programmes et les résultats obtenus par ceux-ci. D'autres sujets comme les thèmes transversaux sont beaucoup moins traités. Les mêmes ordres de grandeur sont observés pour les recommandations. Ceci ressort de la figure ci-dessous qui montre pour 8 thèmes spécifiques le pourcentage de rapports qui fournissent des conclusions qui s'y rapportent et le pourcentage de rapports qui fournissent des recommandations sur ces sujets.

Figure 6 - Thématiques traitées par les conclusions et recommandations des rapports



3.5.2 Conclusions récurrentes

Pertinence des programmes

La plupart des rapports d'évaluation concluent que les programmes financés **répondaient bien à des besoins**. Ils soulignent que les programmes étaient en ligne avec les attentes des bénéficiaires et les besoins généralement bien identifiés. Les réserves émises à ce sujet sont plutôt rares, mais certains rapports concluent que les projets auraient pu être mieux inscrits dans le contexte ou encore que les besoins ont évolué.

Dans cette même veine, plusieurs rapports mettent en exergue la **cohérence avec les stratégies nationales**, tandis que très peu émettent des avis critiques sur ce point.

Par contre, la **conception des stratégies** semble avoir posé problème dans de nombreux cas. Plusieurs rapports soulignent des lacunes en termes de définition du public cible, un manque de priorisation et des activités nombreuses et dispersées. Certains évoquent dans cette optique une absence de cadre stratégique au programme, conçu comme une juxtaposition de projets. De même, des problèmes de décalage entre les objectifs poursuivis et les projets, de structuration du cadre logique et de cohérence entre les différentes mesures sont mentionnés. Il y a toutefois aussi des rapports qui se montrent beaucoup plus positifs sur ces questions et soulignent une très bonne cohérence interne des programmes.

Efficacité et impact

Plusieurs rapports émettent un avis critique sur **les systèmes mis en place pour mesurer les résultats**. Ils soulignent notamment l'absence de système de suivi (monitoring) au niveau de l'action évaluée ou encore des problèmes avec les indicateurs (par exemple parce qu'ils étaient centrés sur la réalisation d'activités plutôt que sur la production de résultats, parce qu'il n'y avait pas de valeur de référence (*baseline*) ou encore parce que les valeurs cibles étaient définies à des niveaux trop ambitieux).

Par contre, et malgré les réserves émises plus haut, il est frappant de constater que de nombreux rapports affirment que les **résultats attendus ont été atteints ou atteints en partie, même si ceci concerne souvent des réalisations, voire des activités. Les rapports restent toutefois vagues sur les questions d'impact**, se limitant la plupart du temps à dire qu'il est trop tôt pour tirer des conclusions en termes d'impact.

Moins d'information est fournie sur les **facteurs favorisant ou entravant l'atteinte des résultats**. On retiendra toutefois les éléments suivants :

- Facteurs favorisant l'atteinte des résultats attendus :
 - Le développement des capacités.
 - L'existence d'un partenaire local avec qui une collaboration de longue date existe et de manière plus générale l'engagement local.

- Facteurs entravant l'atteinte des résultats attendus :
 - Les problèmes de communication.
 - La rotation du personnel.
 - Faiblesses dans les procédures de sélection des destinataires d'une intervention.

Ces éléments restent assez généraux. Ceci ne signifie pas pour autant que les rapports n'en disent pas plus sur des facteurs expliquant l'atteinte ou non les résultats, mais lorsqu'ils le font les éléments fournis sont souvent spécifiques aux programmes (par exemple un outil pédagogique qui était bon, la qualité d'un formateur, etc.). Ceci n'est pas pour autant un problème

Durabilité

Les rapports ne se prononcent pas tellement sur la durabilité des résultats, mais lorsqu'ils le font, c'est généralement en termes négatifs, même s'il y a des exceptions sur ce point. Ils soulignent notamment les risques financiers encourus par certains ANG, lorsque ceux-ci sont dépendants d'un seul bailleur. D'autres mettent en doute la capacité d'autofinancement du projet, voir les capacités des contreparties nationales. Plusieurs rapports soulignent l'absence d'une stratégie de sortie comme risque majeur.

Coordination / synergies

Une bonne coordination est soulignée dans plusieurs rapports. Sur ce point, les rapports font référence à la coordination avec les acteurs locaux ou, par exemple, avec les universités locales. Certains rapports mentionnent, par contre, un manque de coordination entre les acteurs du Nord et du Sud.

Efficiencia

Des exemples de bonne efficacité sont fournis dans certains rapports, tandis que d'autres soulignent pas mal de difficultés et en particulier des retards sur ce point. La lourdeur des procédures (notamment de la DGD) a également été soulignée, mais seulement dans quelques cas et pas suffisamment pour considérer ceci comme une conclusion récurrente.

Thématiques transversales

Les rapports ne se prononcent pas tellement au sujet des thématiques transversales. Lorsqu'ils y font référence, c'est avant tout pour souligner que les problématiques du genre et de l'environnement ont été prises en considération.

3.5.3 Recommandations récurrentes

Pertinence des programmes

Il est difficile de détecter des tendances générales en matière de recommandations relatives à la question de la **réponse aux besoins**. A titre d'exemple, certains rapports préconisent d'accorder une attention explicite aux besoins de groupes spécifiquement vulnérables, d'autres d'éviter d'imposer des thèmes aux organisations partenaires, d'autres encore, de satisfaire le besoin « d'empowerment » des publics cibles. Certains font des propositions concrètes en termes d'objectifs à atteindre par de futurs programmes.

Une recommandation récurrente est de **mieux inscrire les projets dans leur contexte**, en tenant compte des spécificités de celui-ci. Ceci inclut aussi le choix du public cible en fonction de l'analyse des besoins, mais de manière générale aussi de clarifier les publics cibles et de les sélectionner de manière plus rigoureuse.

Les rapports soulignent aussi l'importance d'**améliorer la qualité du cadre logique**, en mettant en cohérence les activités à mettre en œuvre, leurs effets attendus, les facteurs externes et les objectifs à poursuivre. Sur ce dernier point, plusieurs rapports soulignent l'importance de clarifier les objectifs de l'ensemble ou de certaines composantes des programmes.

Certains précisent aussi qu'il faut **vérifier régulièrement la pertinence de la logique d'intervention** et la mettre à jour en fonction de l'évolution des besoins. Il est aussi plusieurs fois préconisé de préciser les modalités de partenariat entre le Nord et le Sud, ou encore le besoin de développer une stratégie pour étendre les projets pilote.

Enfin, plusieurs rapports plaident en faveur d'une **meilleure implication des partenaires locaux** et une meilleure communication de la stratégie auprès des partenaires régionaux.

Efficacité et impact

De manière générale, beaucoup de rapports soulignent l'importance d'**améliorer les systèmes de suivi et évaluation (M&E)**. Différents éléments sont mis en exergue de ce point de vue :

- Définir et hiérarchiser clairement les résultats attendus ;
- Concernant les indicateurs :
 - Définir un nombre limité d'indicateurs (quantitatifs et qualitatifs) ;
 - Définir des indicateurs qui portent sur les résultats et pas uniquement sur les activités ou les 'outputs' ;
 - Définir des valeurs de référence et des valeurs cibles ;
- Définir des systèmes de suivi de programme qui sont en lien avec les instruments utilisés par les partenaires ;
- Améliorer les capacités des équipes locales en matière de M&E, prévoir suffisamment de moyens de ce point de vue et trouver des incitations pour que les personnes rapportent plus régulièrement sur les données ;
- Améliorer le rapportage.

Certains rapports recommandent aussi de **renforcer l'appropriation des résultats** de l'évaluation.

Enfin, les rapports font peu de recommandations permettant d'améliorer l'atteinte des résultats attendus. Lorsque de telles recommandations sont proposées, elles sont très spécifiques aux programmes analysés et donc pas généralisables. Certains soulignent toutefois l'importance de développer les bonnes pratiques ou encore d'augmenter les capacités des organisations bénéficiaires.

Durabilité

En termes de durabilité, les rapports préconisent surtout d'**assurer un financement continu**, par exemple en plaidant en faveur d'un programme additionnel, en soulignant l'importance de diversifier les sources de financement, ou encore en recommandant de demander un financement à l'État. Certains mettent également en exergue l'importance du développement des capacités des organisations.

Coordination / synergie

De nombreux rapports font des recommandations en termes de renforcement de la collaboration avec différentes parties prenantes (autres bailleurs, autorités nationales, entre acteurs du Nord et du Sud), mais sans être très concrets sur les formes que devrait prendre un tel renforcement.

Efficiencia et thématiques transversales

Les rapports se prononcent peu sur la question de savoir comment améliorer l'efficacité. De même ils disent peu sur les thématiques transversales. Une recommandation concerne la création d'un mécanisme permettant l'intégration des thèmes transversaux.

3.6 Utilisation par les ANG – Q6

Comment les ANG utilisent-ils les évaluations?

Cette question porte sur l'utilisation qui est faite par les ANG des rapports d'évaluation examinés. La réponse est structurée en fonction des trois principaux types de finalités qui peuvent être distingués, sans être exclusifs l'un de l'autre: (i) capitaliser sur les connaissances et partager les leçons apprises ; (ii) rendre compte ; et (iii) aider à la décision.⁵⁸

Q6 - Utilisation par les ANG

Résumé de la réponse

La plupart des ANG ont indiqué considérer comme prioritaires pour l'évaluation chacun des trois grands types d'objectifs (capitalisation sur les connaissances et partage de leçons, reddition de compte, et aide à la décision). En pratique, une grande majorité s'est approprié l'évaluation comme un outil d'aide à la décision, pour disposer d'une base solide pour revoir l'approche du programme en cours ou pour préparer le suivant. Pas plus de la moitié a utilisé l'évaluation en pratique pour capitaliser sur les connaissances ou pour rendre compte. Les leçons des évaluations ont en fait rarement été transférées au-delà du cadre de chaque programme. Certains acteurs considéraient par ailleurs que la reddition de compte n'était pas l'objectif prioritaire de l'évaluation mais bien celui du rapportage annuel à la DGD (ce qui était conforme à l'esprit du cadre réglementaire pour les « autres partenaires »).

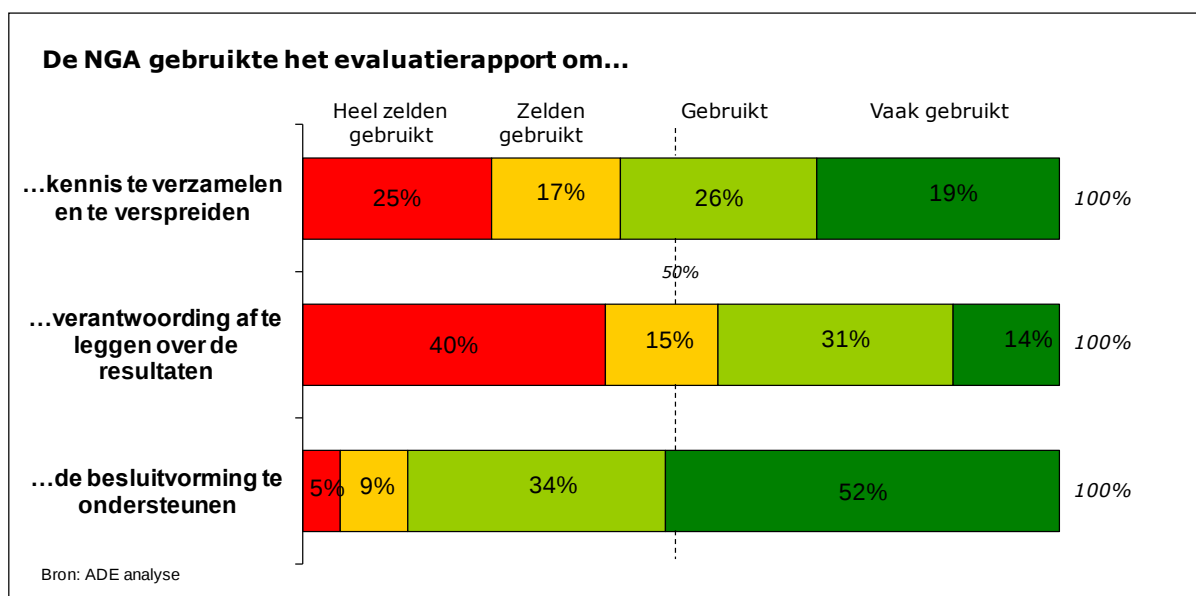
Une grande majorité des ANG ont indiqué considérer comme prioritaires pour l'évaluation chacun des trois types d'objectifs (apprentissage, reddition de compte et prise de décision). Ceci dit, les ONG donnent plus souvent priorité à l'apprentissage. Pratiquement toutes les ONG (91%) ont répondu à l'enquête qu'un objectif prioritaire de l'évaluation est d'accumuler les connaissances et de transférer les leçons apprises. Les trois quarts d'entre elles ont répondu que rendre compte de l'atteinte des résultats attendus était un objectif prioritaire (74%), tout comme aider à la prise de décision (76%). Les 'autres partenaires' ont donné le même ordre de priorité aux trois objectifs (91% à chaque fois), tout comme les universités et institutions scientifiques sur base de leurs réponses à l'enquête et des entretiens.⁵⁹

En pratique, une grande majorité des ANG a utilisé les évaluations sélectionnées comme un outil d'aide à la décision. La moitié des ANG a également utilisé ces évaluations pour capitaliser sur les connaissances et pour rendre compte, comme indiqué dans la figure ci-dessous.⁶⁰

⁵⁸ Cette distinction est faite par exemple dans la méthodologie de l'Unité d'évaluation des relations externes de l'UE (au sein de EuropeAid) ou dans la publication d'ITECO, Antipodes, décembre 2010, *Évaluer l'évaluation*, p9.

⁵⁹ Les réponses à l'enquête pour les universités et institutions scientifiques ne sont pas fiables dans la mesure où 2 sur les 6 ont (étonnamment) répondu « NA » (non applicable) à cette question.

⁶⁰ Ces chiffres proviennent de la synthèse des Fiches acteurs, qui sont le résultat de l'analyse des réponses et documents envoyés par les ANG suite aux demandes individuelles par email, des réponses aux enquêtes, et des entretiens et focus groups.

Figure 7 - Utilisation des rapports par les ANG⁶¹

Des détails sur chacun de ces trois types d'utilisation sont fournis ci-dessous.

3.6.1 Capitalisation sur les connaissances et partage de leçons (Q6.1)

Si tous les ANG considéraient l'accumulation des connaissances comme un objectif prioritaire, ceci n'a vraiment eu lieu que dans une bonne moitié des évaluations.

- Pratiquement tous les ANG (89%) ont répondu à l'enquête qu'un objectif prioritaire de l'évaluation était d'accumuler les connaissances et de transférer les leçons apprises (cf. supra).
- Une bonne moitié des rapports ou termes de référence des évaluations (59%) indiquaient que le *processus d'évaluation* prévoyait l'identification des bonnes pratiques et de leçons par les parties au projet (i.e. lors de l'organisation d'ateliers participatifs, de « focus groups », de formations pour les parties prenantes, réunions de débriefing, etc.).⁶² Plusieurs représentants d'ANG ont par ailleurs indiqué avoir beaucoup appris eux-mêmes sur l'évaluation en contribuant à ces évaluations financées par la DGD, en particulier des 'autres partenaires' (pour lesquels la DGD, voire le SES, avait un rôle plus important dans un comité de concertation).
- L'étude des *rapports d'évaluation* montre que les bonnes pratiques et les leçons apprises étaient soulignées dans 56% des rapports, que ce soit au travers d'une section spécifique, du chapitre de conclusions ou d'une autre façon. Un document séparé voire une brochure ont par exemple parfois même été rédigés.
- Une *restitution des résultats* de l'évaluation après remise du rapport final a eu lieu (ou du moins était prévue dans les TdR), au Nord et/ou au Sud, dans une bonne moitié des évaluations également (55%).

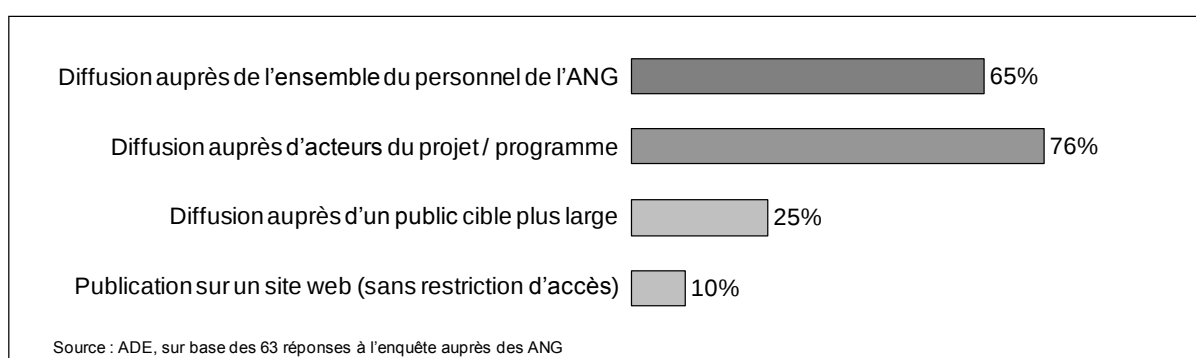
⁶¹ Les ordres de grandeur pour les trois types d'ANG (ONG ; universités et institutions scientifiques ; et autres partenaires) sont similaires.

⁶² L'évaluation a été conduite selon une approche participative dans deux tiers des cas (64%), comme décrit dans la section 4.2.3 relative au processus d'évaluation.

Les leçons d'une évaluation ont régulièrement été partagées en interne et auprès des parties prenantes directes, mais pas systématiquement, et rarement au-delà. Une initiative interne au secteur des ONG n'a pas eu le succès escompté.

- Une grande majorité des ANG n'ont pas cherché à diffuser activement les leçons apprises des évaluations (ni les rapports d'évaluation en tant que tels), au-delà des collaborateurs internes et des parties prenantes directes aux activités. Seulement 6 rapports étaient librement accessibles à toute personne intéressée via le site web de l'ANG (3 rapports d'ONG, 2 d'universités et 1 d'un autre partenaire) – voir la figure 8 ci-dessous. Certains acteurs (27%) ont néanmoins diffusé plus largement une partie du rapport, par exemple le résumé. Ceci fut par exemple le cas de l'évaluation de la VVSG⁶³, dont les principaux éléments ont été présentés à un congrès de pouvoirs locaux.

Figure 8 - Diffusion des rapports d'évaluation⁶⁴



- Les fédérations d'ONG avaient chargé ATOL et le COTA de créer une **base de données des rapports d'évaluation** d'ONG relatifs à la période 2002-2007, pour les rendre accessibles aux autres ONG. Ceci était une des recommandations de l'évaluation du « 1% » (cf. section 3.1.1). L'initiative n'a toutefois pas réellement été un succès d'après plusieurs sources, en raison notamment de la réticence de nombreuses ONG à partager leurs rapports d'évaluation et à la lourdeur du système créé (collecte, encodage, accès, attrait, etc.). L'idée a été partiellement reprise dans une nouvelle base de données appelée Viungo, lancée en 2011. Également accessible uniquement aux ONG belges, elle ne permet pas de partager des leçons avec d'autres acteurs ou personnes intéressées.

⁶³ Vereniging van Vlaamse Steden en Gemeenten

⁶⁴ Les ordres de grandeur pour les trois types d'ANG (ONG ; universités et institutions scientifiques ; et autres partenaires) sont similaires.

3.6.2 Reddition de compte (Q6.1)⁶⁵

Si la plupart des ANG considéraient la reddition de compte comme un objectif prioritaire pour l'évaluation, ceci ne s'est vérifié que dans moins de la moitié des évaluations. Certains acteurs considéraient par ailleurs que la reddition de compte n'était pas l'objectif prioritaire de l'évaluation mais bien celui du rapportage annuel à la DGD, ce qui était conforme à l'esprit du cadre réglementaire pour les « autres partenaires ».

- Les trois quarts des **ONG** considéraient que la reddition de compte sur les résultats atteints était un objectif prioritaire pour l'évaluation. Plusieurs interlocuteurs ont indiqué qu'ils considéraient cependant la reddition de compte comme l'objet du rapportage annuel et final par les acteurs à la DGD. Si le cadre réglementaire pour les ONG ne donnait pas d'indication sur le sujet, celui pour les 'autres partenaires' bien. Il allait en effet dans ce sens, citant l'objectif d'« amélioration prospective » ; la reddition de compte sur les résultats et les activités était quant à lui dévolu au rapportage annuel (cf. section 3.1.1). Étonnamment, dans cette perspective réglementaire, 91% des « **autres partenaires** » considéraient tout de même la reddition de compte comme un objectif prioritaire pour leur évaluation.⁶⁶
- La grande majorité des **universités et institutions scientifiques** considérait que rendre compte des résultats atteints était un objectif prioritaire pour l'évaluation. Plusieurs interlocuteurs nous ont en effet expliqué trouver cela naturel vu le financement public dont ils ont pu bénéficier, et dont ils dépendent largement (voir section 3.1.1). Tout comme les ONG et d'autres partenaires, ils insistent également sur l'importance de rendre compte des résultats atteints aux partenaires locaux et, en particulier, aux bénéficiaires, au-delà des donateurs (DGD, autres bailleurs, grand public, etc.).
- L'analyse montre qu'**en pratique**, moins de la moitié des ANG ont vraiment utilisé l'évaluation pour la reddition de compte sur les résultats (voir figure 7 ci-dessus). Ce constat se base sur une analyse des objectifs assignés à l'évaluation dans les TdR, des questions d'évaluation, des réponses aux enquêtes, du choix des activités à évaluer (fréquemment plus orienté vers une amélioration des activités qui seront continuées dans la suite du programme que vers la reddition de compte sur les résultats atteints de toutes les activités en ce compris des activités qui seront arrêtées), etc. Les conclusions d'un grand nombre de rapports (42%) ne donnaient pas de jugement d'ensemble clair sur le niveau d'atteinte des objectifs du programme. De nombreux ANG n'avaient par ailleurs pas envoyé leur rapport d'évaluation à la DGD ou du moins pas spontanément (voir Q7). Un quart des ANG (27%) indique dans l'enquête n'avoir pas ou peu utilisé le rapport d'évaluation examiné dans le dialogue politique avec la DGD, avec des différences importantes par type d'acteur : 29% des ONG, 50% (!) des universités et institutions scientifiques, mais seulement 9% des 'autres partenaires'.

⁶⁵ 'Accountability' en anglais - également appelée en français rendu de comptes, redbilité ou imputabilité.

⁶⁶ Ceci est peut-être dû en partie aux instructions additionnelles données par les gestionnaires de la DGD, comme indiqué par certains interlocuteurs, la DGD faisant d'office partie du comité de concertation chargé d'approuver les termes de référence (voir section 4.1.1)

3.6.3 Prise de décision (Q6.3)

Les ANG se sont largement approprié l'évaluation pour disposer d'une base solide pour revoir l'approche du programme en cours ou pour préparer le suivant. Ceci est conforme à l'esprit du cadre réglementaire pour les 'autres partenaires'. Peu ont toutefois intégré l'évaluation de façon systématique dans leur processus de gestion (réponse managériale, plan d'action subséquent, etc.).

- Les trois quarts (76%) des ONG, la plupart des universités et institutions scientifiques et même 91% des 'autres partenaires' considéraient l'évaluation comme un outil d'aide à la prise de décision.
- En pratique, l'analyse montre que les ANG ont très largement (86%) utilisé l'évaluation à des fins d'aide à la prise de décision, y compris les ONG (86%). Ceci ressort des termes de référence des évaluations, des réponses aux enquêtes, des entretiens et focus groups, etc. De nombreux ANG ont témoigné s'être appropriés l'évaluation pour disposer d'informations utiles pour prendre des décisions par rapport à la poursuite du programme en cours (corriger, réviser, réallouer, recentrer, etc.) ou pour préparer le programme suivant.
- Une moitié des ANG (47%) a pris une position claire sur les recommandations de l'évaluation dans la foulée du rapport d'évaluation, par exemple dans une « management response » (38% n'ont rien produit de tel et 14% de façon peu systématique). Un cinquième des ANG (22%) a par ailleurs rédigé un plan d'actions crédible sur base des différentes recommandations qu'elle jugeait utiles, p.ex. actions à prendre, échéances, responsabilités (45% n'ont rien établi et 18% de façon peu systématique). Certains ANG décrivaient par ailleurs dans des documents internes ou externes, ou dans leur documents de dialogue politique avec la DGD, comment ils comptaient tenir compte de l'une ou l'autre des recommandations.
- Il ressort également des entretiens et commentaires aux enquêtes que quand de nombreux ANG indiquaient qu'ils voulaient que l'évaluation leur permît « d'apprendre et de tirer les leçons », dans une perspective d'« **organisation apprenante** », ils référaient souvent à tenir compte des leçons du passé dans les décisions pour le futur, plutôt qu'à accumuler et partager les connaissances en tant que tel.

Encadré 8 – Moment de l'évaluation

De nombreux ANG ont témoigné de l'importance du choix du moment approprié pour conduire une évaluation afin qu'elle leur fût directement utile. Pour les évaluations visant la *reddition de compte*, celui-ci est typiquement après la fin de la mise en œuvre des activités, voire une ou quelques années par après pour avoir le recul nécessaire pour apprécier la durabilité des effets et l'impact. Le moment le plus utile pour des évaluations visant à *aider à la décision* (par les ANG ou par la DGD) est à déterminer dans chaque cas en fonction de la date envisagée pour la prise de décision, tout en veillant à disposer de suffisamment d'informations sur les résultats obtenus. Le meilleur moment de l'évaluation pour celles visant la *capitalisation et le partage des connaissances* est souvent moins critique et dépend surtout du contexte et du type de leçons à tirer.

Deux difficultés du système de cofinancement pour la période 2008-2013 (avec généralement un cadre stratégique sur 6 ans couvrant deux programmes successifs de 3 ans) étaient la confusion sur ces types d'objectifs et le fait de réaliser une évaluation à mi-parcours et une finale sur cette période⁶⁷. Pour celles qui visaient principalement (de façon implicite) à la prise de décision par les ANG, il n'était pas évident dans un tel laps de temps de trouver un moment permettant d'avoir le recul nécessaire pour pouvoir observer des résultats et pour pouvoir encore tenir compte des recommandations dans la préparation du programme suivant. Cela a souvent conduit à forcer des calendriers serrés pour l'évaluation (p.ex. en début d'année n+2 avant de soumettre la proposition pour le second programme), ainsi qu'à trouver des arbitrages inadéquats. Un calendrier trop serré a eu un impact sur la qualité de l'évaluation dans plusieurs cas (voir section 3.2.1).

Plusieurs gestionnaires de la DGD ont indiqué qu'il était important également pour eux de disposer des évaluations à un bon moment, qui était pour eux celui leur permettant de les utiliser dans leur appréciation des propositions de programme suivant (cf. Q7).

⁶⁷ Notons que la distinction entre évaluations 'à mi-parcours' et 'finale' s'est souvent avérée trompeuse dans cette étude et a donc généralement été évitée. Des évaluations dites à mi-parcours (par rapport au cadre stratégique sur 6 ans) pouvaient en effet être une appréciation finale d'un programme de 3 ans. A l'inverse, une évaluation dite finale du premier programme de 3 ans était en fait à mi-parcours de la période totale de 6 ans.

3.7 Utilisation par la DGD – Q7

Comment la DGD utilise-t-elle les évaluations ?

Cette question porte sur l'utilisation qui est faite des rapports d'évaluation des programmes d'ANG par la DGD, en particulier par les gestionnaires des différents départements de l'(ex) D3 de la DGD en charge du cofinancement des 'Programmes non gouvernementaux'.⁶⁸ La réponse à la question examine tout d'abord les directives de la DGD sur l'utilisation par son personnel des rapports d'évaluation, ainsi que la mesure dans laquelle la DGD a reçu les rapports d'évaluation des ANG. Elle étudie ensuite la façon dont la DGD a utilisé ces rapports en pratique et la façon dont celle-ci s'est organisée à cette fin. Les pistes d'amélioration pour l'utilisation des évaluations sont présentées dans le chapitre sur les recommandations.

Q7 - Utilisation par la DGD

Résumé de la réponse

Les évaluations ont été utilisées par la DGD dans le dialogue politique avec les ONG et en particulier pour l'appréciation de nouvelles phases de programmes, même si de façon très variable. Ceci fut nettement moins le cas avec les autres types d'acteurs.

Une grande majorité des gestionnaires de la DGD considérait que les évaluations externes avaient une valeur ajoutée par rapport au rapportage annuel par les ANG, fournissant une information complémentaire, (plus) indépendante et allant au-delà de la gestion du programme.

La DGD ne disposait toutefois pas d'une part significative des rapports d'évaluation, en particulier ceux d'ONG, et n'avait reçu nombre d'autres qu'après une demande expresse. Ceci résultait en particulier du fait que le cadre réglementaire ne prévoyait pas la transmission des rapports d'évaluation à la DGD.

La DGD n'avait pas non plus précisé comment elle comptait utiliser ces évaluations. Elle disposait en fait d'une faible capacité et organisation internes en évaluation, limitant ainsi la compréhension des enjeux, l'utilisation des rapports et la diffusion des enseignements.

⁶⁸ L'ex-D3 'Programmes non gouvernementaux' était composée principalement de la D3.1 'ONG', de la D3.2 'Universités et Institutions scientifiques' et de la D3.3 « Autres partenaires ». Il y avait par ailleurs une D3.4 'Programmation budgétaire et contrôle financier pour les programmes non gouvernementaux'. Ce département D3 est maintenant dénommé 'Société civile' et est subdivisé en fonction des régions d'intervention.

3.7.1 Directives sur l'utilisation et Réception des rapports (Q7.1 et Q7.2)

Le cadre réglementaire pour chacun des types d'ANG ne précise pas que les ANG doivent envoyer les rapports d'évaluation à la DGD (cf. Q1). Il n'indique pas non plus précisément comment la DGD compte les utiliser. Il est simplement demandé aux « autres partenaires » d'envoyer des « synthèses » des rapports d'évaluation en max. 5 pages accompagnées des TdR et des rapports de mission, dans le mois qui suit le dépôt du rapport final.⁶⁹ Par ailleurs, pour les partenaires universitaires, la DGD étant dans le comité de concertation « en charge du suivi et de l'évaluation », il peut être raisonnablement attendu qu'elle reçoive les rapports d'évaluation. Il n'y a pas de précisions particulières pour les ONG, si ce n'est indirectement pour l'appréciation de la proposition de programme suivant.⁷⁰ Plus de la moitié des gestionnaires de la DGD ont toutefois donné des instructions additionnelles au cadre réglementaire sur l'envoi du rapport à la DGD (57%).

De nombreux interlocuteurs, ainsi que l'évaluation récente du rapportage sur les résultats⁷¹, mentionnent à ce propos le **déséquilibre flagrant entre le travail attentif et minutieux d'appréciation par la DGD des propositions de programme d'ANG, et l'absence d'indications relatives à la prise en compte des évaluations et des informations sur l'atteinte des résultats.** Un gestionnaire a expliqué que la DGD avait beaucoup misé sur l'appréciation des programmes en phase de préfinancement, afin de disposer de bases solides et de pouvoir lâcher prise par la suite. C'était également lié à une culture organisationnelle plus portée sur les intrants que sur les résultats.

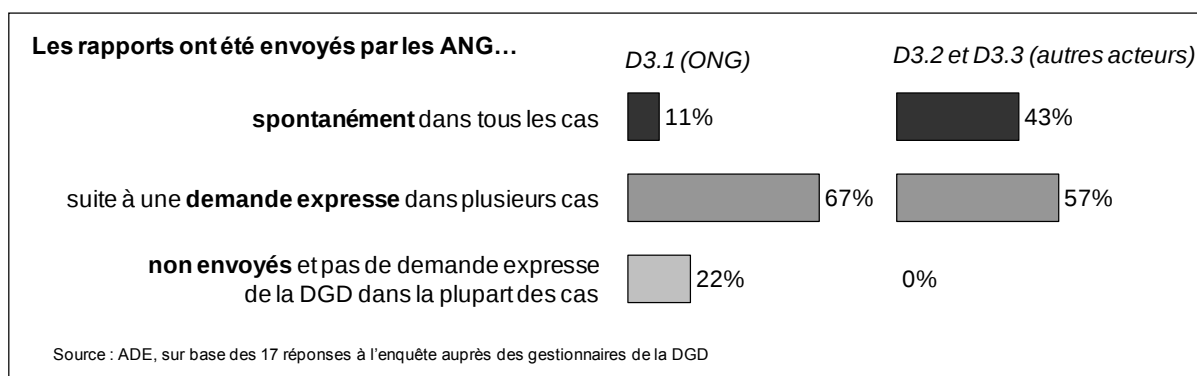
La DGD ne disposait ainsi pas d'une part significative de rapports d'évaluation, en particulier d'ONG, et n'avait reçu nombre d'autres qu'après une demande expresse.

- Trois des dix (3/10) gestionnaires de programmes d'ONG à la DGD ayant répondu à l'enquête ont indiqué n'avoir eu que peu souvent accès au rapport final (et annexes) des évaluations ; 5/10 ont répondu qu'ils l'ont eu dans la plupart des cas et (seulement) 2/10 systématiquement. Ces chiffres sont meilleurs pour les autres types de partenaires (4/7 'systématiquement', 2/7 'dans la plupart des cas et 1/7 'rarement'). L'absence de rapports à la DGD pour des nombreux ANG avait par exemple posé problème pour déterminer l'échantillon pour cette méta-évaluation. Dans quelques cas il n'y avait en fait pas eu d'évaluation à proprement parler (voir section 2.2).
- Les rapports reçus l'ont souvent été suite à une demande expresse de la DGD, pour les ONG en particulier (voir figure ci-dessous). Cela a également été le cas pour d'autres types d'acteurs. Un gestionnaire de l'ex-D3.2 à la DGD a par exemple indiqué avoir « reçu du feedback général en comité de concertation, mais dû insister pour obtenir le rapport d'évaluation ».

⁶⁹ Voir Annexe 2.

⁷⁰ La DGD appréciait les propositions de nouveaux programme des ANG sur base de 82 critères, dont le suivant : « *Les leçons apprises du programme précédent ont-elles été intégrées?* » (critère #44 de la Fiche d'appréciation – cf. Annexe 2). De telles leçons pouvaient venir, par exemple, de l'évaluation externe.

⁷¹ Service de l'Évaluation spéciale, 2012. *Évaluation du rapportage des résultats de la Direction générale de la Coopération au Développement*, SPF Affaires Étrangères, Commerce extérieur et Coopération au développement.

Figure 9 - Réception des rapports d'évaluation

3.7.2 Utilisation des évaluations par la DGD (Q7.1 et Q7.2)

Une grande majorité des gestionnaires de la DGD ont indiqué que les évaluations externes avaient une valeur ajoutée par rapport au rapportage annuel par les ANG et qu'ils ont vérifié la prise en compte par les ANG des leçons et recommandations. Les évaluations ont également été largement utilisées dans le dialogue politique avec les ONG, quoique nettement moins avec les autres types d'acteurs.

- 82% des gestionnaires de la DGD ayant répondu à l'enquête ont indiqué que les **rapports d'évaluation avaient une valeur ajoutée** (35%) voire une grande valeur ajoutée (47%) dans leur travail de suivi des programmes des ANG, **en plus du rapportage narratif et financier** (seuls 18% ont répondu qu'ils avaient peu souvent de valeur ajoutée). Les évaluations avaient selon eux pour principal avantage de fournir une information complémentaire, (plus) indépendante et potentiellement avec un autre point de vue. Certains ont également indiqué que l'évaluation permettait de voir au-delà de la gestion du programme, dans une « optique de compréhension (pas qu'administrative), avec un jugement plus tranché sur l'atteinte des objectifs et sur les réflexions. ». « *Quand je reçois un 'rapport de programme' d'ONG, témoigne une gestionnaire, souvent je commence par regarder le rapport d'évaluation, dans la mesure où il est sensé être général et externe. Il facilite notre compréhension du 'rapport de programme'. Mais il peut être de type et de qualité très variables...* ». Des représentants d'ONG ont soulevé que la place donnée à l'évaluation dans le suivi dépendait en fait souvent d'un gestionnaire à l'autre.
- 82% également des gestionnaires de la DGD ont indiqué avoir examiné la **prise en compte par les ANG des leçons apprises et recommandations** du rapport d'évaluation (réponse managériale formelle, développement d'un plan d'actions, etc.), systématiquement (47%) ou dans la plupart des cas (35%). Seuls 18% des gestionnaires ont répondu ne l'avoir fait que peu souvent. Certains gestionnaires ont indiqué qu'ils en tenaient compte lors du rapportage annuel (notamment au travers de la 'note de synthèse'), lors de missions de terrain, ou, davantage, lors de l'appréciation de la proposition de programme suivant, pour autant qu'ils fussent disponibles à ce moment-là (voir encadré 8 dans la section 3.6.3). Quelques cas ont été rapportés d'après discussions entre la DGD et des ANG, notamment lorsque l'ANG ne partageait pas certaines recommandations de l'évaluation. Ceci était par ailleurs compliqué par l'absence de précisions dans le cadre réglementaire sur l'utilisation à faire par la DGD (et les ANG) des leçons et recommandations des évaluations.

- Une différence importante est par contre observée entre les gestionnaires de programmes d'ONG et les autres sur **l'utilisation des rapports d'évaluation dans le cadre du dialogue politique** avec ces partenaires. 90% des gestionnaires de la D3.1 (ONG) indiquent les avoir utilisés souvent (80%) ou systématiquement (10%), contre 43% des gestionnaires de la D3.2 et D3.3 (pour lesquels un autre 43% a répondu les avoir utilisés « peu souvent »). Certains interlocuteurs ont indiqué à ce propos qu'il n'y avait en fait pas de 'dialogue politique' avec les universités ni les institutions scientifiques, même si une interaction régulière au sein des comités de concertation ou de pilotage, et que l'implication dans l'évaluation consistait en fait principalement à surveiller la procédure de marché public.

Encadré 9 – Relation entre la DGD et les ANG

La relation entre la DGD et les partenaires universitaires et scientifiques était reconnue par pratiquement toutes les parties comme généralement basée sur la confiance mutuelle.

L'interaction entre la DGD et les ONG et « autres partenaires » se faisait toutefois souvent plus dans une relation de contrôle, comme indiqué par des représentants des diverses parties. Les différentes interprétations du cadre réglementaire n'ont pas facilité les choses, tout comme les capacités parfois limitées des parties en matière d'évaluation, tant au sein de la DGD (cf. infra) que des ANG (cf. section 3.1.3).

Certaines pratiques positives ont toutefois été relevées par plusieurs interlocuteurs, comme la participation de certains fonctionnaires de la DGD ou du SES à des formations organisées par les fédérations d'ONG, ou l'invitation d'un gestionnaire de la DGD par une ONG à la présentation de son évaluation.

Enfin, notons qu'un bureau d'études avait développé une pratique intéressante dans les évaluations qu'il avait conduites à la demande des ANG: il y formulait des recommandations à l'adresse de la DGD également. Ceci a permis de faire passer des messages directement aux gestionnaires et d'en favoriser ainsi la compréhension et la prise en compte par la DGD.

3.7.3 Organisation interne (Q7.3)

La DGD ne disposait que d'une faible capacité en évaluation, limitant la compréhension des enjeux et l'utilisation des rapports d'évaluation.

- La responsabilité du suivi d'un programme d'ANG incombait principalement à un gestionnaire de l'ex-direction D3 'Programmes non gouvernementaux'. Plusieurs ont indiqué en entretien ou au travers des enquêtes que peu d'entre eux avaient une réelle connaissance des concepts de l'évaluation, tout comme l'ensemble du personnel de la DGD au siège ou dans les pays partenaires. Certains ont eu l'opportunité de suivre une courte formation en évaluation, donnée par exemple avec l'appui de l'unité interne d'évaluation existant à l'époque, mais ce ne fut pas systématiquement le cas. Plusieurs gestionnaires et d'autres interlocuteurs ont indiqué que leur peu de maîtrise des concepts de l'évaluation a affecté leur compréhension des rapports et leur capacité à les interpréter et les utiliser activement. Cela a également entraîné de la résistance par rapport à des pratiques innovantes, par manque de capacité de jugement sur l'adéquation des outils proposés (cf. section 3.2.2).
- L'évaluation du rapportage des résultats de la DGD mentionne à ce propos que « La DGD ne possède toutefois pas la capacité d'évaluation lui permettant de soutenir les directions. Une telle capacité est nécessaire afin de parvenir à une planification

efficace de l'évaluation et à une due exécution de qualité de ladite évaluation. »⁷² La DGD avait en fait créé une unité interne en charge de l'évaluation dans les années 1990. Après un passage à vide entre 1999 et 2003 au moment de la création de la DGCI (ex-DGD) et de la CTB⁷³, d'après des fonctionnaires en place à l'époque, une unité interne d'évaluation avait été relancée en 2003 pour répondre aux besoins internes en évaluation. Ceci se faisait en complément du SES, un service externe rapportant directement au Ministre de la Coopération au Développement, avec ses propres budget et programme d'évaluation. La fonction d'évaluation interne à la DGD a été dissoute en 2010, notamment parce que son champ d'activités était très similaire en pratique à celui du SES ; plusieurs de ses fonctionnaires ont été intégrés au SES. Un réseau relatif à la gestion axée sur les résultats avait en outre été créé au sein de la DGD, qui permettait par exemple de donner des formations touchant à l'évaluation ; celui-ci a été dissout récemment, comme les autres réseaux à la DGD. Il reste à la DGD un petit budget pour réaliser des évaluations ad hoc. Elle a également créé un service interne en charge de la qualité et des résultats, mais sans mandat en évaluation. La DGD ne dispose ainsi plus à l'heure actuelle d'un service interne lui permettant structurellement de répondre à des demandes de sa direction et de son personnel en évaluation.⁷⁴

- Plusieurs gestionnaires et interlocuteurs ont indiqué que les gestionnaires n'examinaient les évaluations souvent que de façon limitée, étant donné le nombre d'autres documents à lire par acteur (et leur taille parfois) et le nombre d'acteurs dont certains avaient la charge, en particulier les gestionnaires d'ONG.⁷⁵ Certains acteurs n'avaient par ailleurs pas fait réaliser d'évaluation globale de leur programme, mais bien une kyrielle d'évaluations d'activités diverses. L'évaluation du rapportage des résultats de la DGD mentionne également à propos des évaluations que « Ces dernières sont régulièrement exécutées, mais la communication à la DGD des informations collectées sur la base de ces évaluations et la capacité d'absorption de ce volume d'informations par la DGD, sont limitées. »⁷⁶
- La rotation des responsabilités a également affecté la connaissance des dossiers et la cohérence des instructions données aux ANG – certains acteurs ont indiqué que jusqu'à 4 gestionnaires différents ont été en charge du suivi de leur programme en quelques années. Le suivi pouvait également dépendre d'un gestionnaire à l'autre. Certains acteurs ont indiqué ne pas avoir eu de dialogue depuis 2 ans.

⁷² Service de l'Évaluation spéciale, 2012. *Évaluation du rapportage des résultats de la Direction générale de la Coopération au Développement*, SPF Affaires Étrangères, Commerce extérieur et Coopération au développement, p56.

⁷³ Agence belge de développement (CTB - d'après la dénomination officielle de Coopération technique belge).

⁷⁴ La capacité en évaluation combinée de la DGD et du SES a par ailleurs diminuée entre 2003 et 2011, tant en termes d'enveloppe budgétaire que de personnes (d'environ 9 fonctionnaires spécialisés en évaluation en 2003 à environ 7 actuellement).

⁷⁵ Il ressort de l'enquête auprès des gestionnaires de programmes d'ONG que 82% de ceux qui ont répondu avaient été impliqués dans le suivi de 4 à 10 programmes d'ONG durant la période 2008-2010 (9% en avaient eu moins de 4 et 9% en avaient eu plus de 10). Ce chiffre était de 33% pour les deux autres services, où près de la moitié des gestionnaires (44%) suivait entre 1 et 3 acteurs.

⁷⁶ Service de l'Évaluation spéciale, 2012, *Ibid.*, p44.

L'organisation interne n'a pas favorisé le partage des informations relatives à l'évaluation au sein de la DGD.

- Moins d'un tiers des gestionnaires de la DGD a répondu positivement à l'enquête à la question si la DGD était « organisée de manière à maximiser l'utilisation des évaluations des programmes des ANG (p.ex. stockage et circulation de l'information, etc.) ». Ceci était particulièrement le cas des gestionnaires de programmes d'ONG, avec 7 sur 10 d'entre eux trouvant que la DGD était mal (2/10) voire très mal (5/10) organisée de ce point de vue.
- Un problème cité de façon récurrente était que les rapports d'évaluation n'étaient pas classés de façon systématique sur le serveur interne, comme c'était pourtant le cas par ANG pour les documents relatifs à l'appréciation des programmes, les procès-verbaux des dialogues politiques, etc. Un système SharePoint interne permettait toutefois de partager les rapports, mais l'utilisation n'en était pas optimale et son usage pas généralisé.
- Plusieurs gestionnaires ont toutefois souligné « l'esprit d'entraide et de communication entre collègues, qui se transmettent facilement des informations et des rapports ». Néanmoins, d'autres ont attiré l'attention sur le fait que chaque gestionnaire avait une grande autonomie dans le suivi de « ses » ANG, chacun l'effectuant en parallèle, avec peu d'échanges sur le contenu ou la méthodologie, et peu d'efforts d'harmonisation aboutis. Ceci a résulté entre autres dans des instructions personnelles variées aux ANG (cf. section 3.1.1).
- Plusieurs gestionnaires ont également indiqué qu'il n'y avait (pratiquement) pas eu jusqu'à présent de capitalisation interne des enseignements et leçons des évaluations des programmes d'ANG, ni en continu ni de façon synthétique.

4. Conclusions

Ce chapitre présente (i) notre appréciation générale de l'évaluation des programmes des acteurs non gouvernementaux cofinancés par la DGD ; (ii) les conclusions générales émergeant de cette méta-évaluation.

Les constats sur lesquels se basent les conclusions sont présentés dans le Chapitre 3.

4.1 Appréciation générale

La coopération non gouvernementale a représenté 901 millions € ou 18% du total de l'aide au développement de la DGD durant la période 2008-2011. La DGD a donné la responsabilité de l'évaluation des programmes cofinancés des acteurs non gouvernementaux à ces mêmes acteurs, sans pour autant l'inscrire dans une véritable politique d'évaluation ni donner de directives claires en la matière.

La DGD n'a en conséquence que peu pu utiliser le potentiel de l'évaluation pour ses propres besoins, que ce soit pour rendre compte sur son aide, pour aider à sa prise de décision, ou pour capitaliser et partager les leçons tirées des programmes. Elle ne disposait par exemple pas d'une part significative des évaluations, suite au fait que le cadre réglementaire ne prévoyait pas que les ANG dussent les lui transmettre.

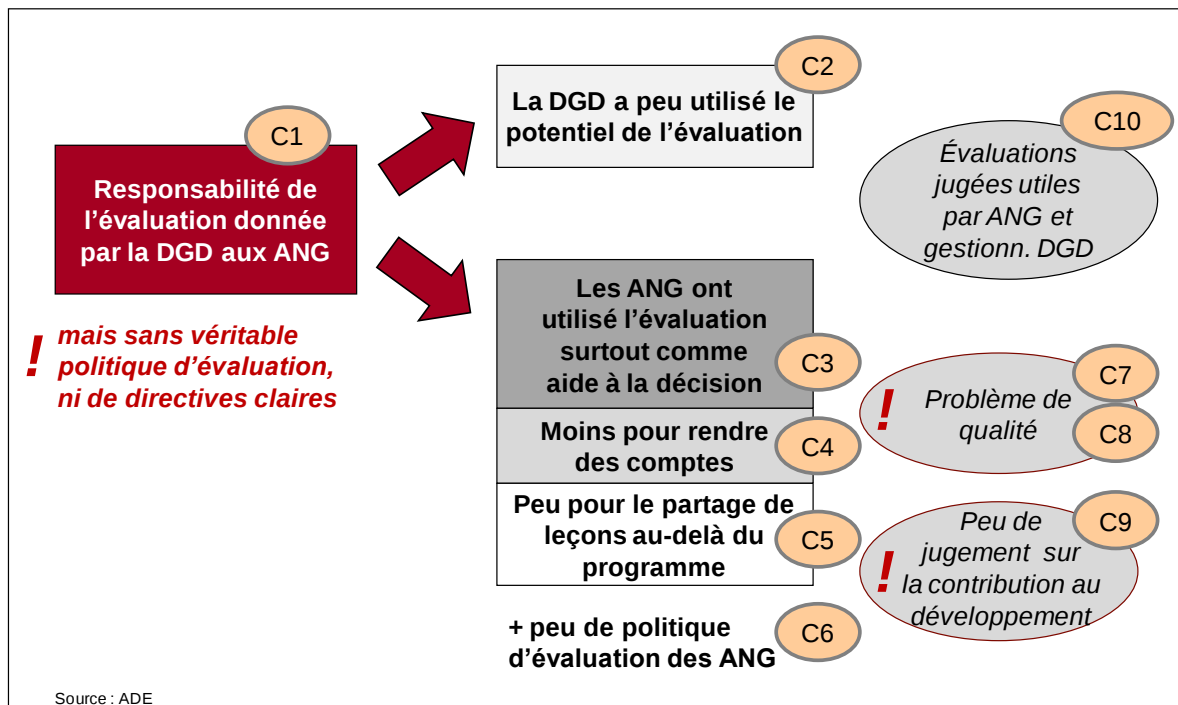
Les ANG ont fait réaliser des évaluations principalement dans l'optique d'améliorer la poursuite de leur programme. Ils l'ont nettement moins utilisée dans une perspective de reddition de compte ou de partage des leçons avec de tierces parties.

La qualité des rapports d'évaluation pose problème. Même si elle est bonne dans l'ensemble pour une grande majorité des programmes d'universités et d'institutions scientifiques, elle est faible pour la moitié des programmes d'ONG et d'autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité. Les progrès des capacités des ANG en évaluation restent par ailleurs modestes.

4.2 Conclusions

Les dix conclusions transversales émergeant de cette étude sont présentées dans la figure ci-dessous et détaillées par la suite.

Figure 10 - Conclusions



Chaque conclusion détaille les constats sur lesquels elle se base, avec indication du numéro de la question d'évaluation (Q1 à Q7) au sein de laquelle ceux-ci sont formulés en plus de détail, dans le Chapitre 3.

Conclusion 1: Absence d'une véritable politique d'évaluation

Pour la période 2008-2011, il n'y avait pas à la DGD de véritable politique d'évaluation fixant des objectifs, définissant des modalités et prévoyant les ressources nécessaires. Celle-ci s'est largement limitée à transférer la responsabilité de l'évaluation aux acteurs.

Basée sur les questions Q1 et Q7

Cette conclusion est basée sur les constats suivants:

- La coopération non gouvernementale a représenté 901 millions € ou 18% du total de l'aide au développement de la DGD durant la période 2008-2011.⁷⁷ La DGD en a donné la responsabilité de l'évaluation aux acteurs non gouvernementaux (ANG) eux-mêmes. Ceci était en lien avec le principe de responsabilisation et de professionnalisation de ces acteurs en matière d'évaluation défini lors de la réforme à la fin des années 1990. Une politique d'évaluation globale n'avait cependant pas été définie, ce qui n'était pas en lien avec la responsabilité de l'évaluation de la politique de coopération au développement donnée à l'administration. (Q1)
- Les directives en matière d'évaluation consistaient en fait en des références plus ou moins explicites dans une série de documents de types divers et variés. Celles-ci laissaient aux ANG une grande liberté en la matière, avec un rôle souvent limité voire inexistant de la DGD, en particulier pour les ONG. Vu qu'elles étaient peu claires ou précises, les gestionnaires de la DGD ont souvent dû donner des clarifications et ont en outre donné des instructions personnelles additionnelles, sur base de leur sensibilité à l'évaluation. (Q1)
- L'absence de véritable politique d'évaluation pour la période d'évaluation 2008-2011 tranchait avec la politique volontariste du « 1% » dans les années ayant suivi la réforme (1% du budget du programme devant être consacré à de l'évaluation), qui avait amorcé un renforcement structurel du secteur des ONG belges en matière de culture d'apprentissage et d'évaluation. (Q1) Cela résulte de l'absence d'une vision claire sur l'évaluation depuis lors, liée entre autres à l'absence d'une culture forte en évaluation au sein de la DGD. Certains acteurs avaient par ailleurs insisté pour obtenir la maîtrise de l'évaluation, afin de pouvoir l'utiliser au mieux en fonction des besoins de chacun. C'était aussi une façon pour la DGD de pallier à la pression croissante sur ses ressources humaines. (Q7)

⁷⁷ Montants et pourcentage calculés sur base des chiffres publiés dans le Rapport annuel 2011 de la DGD. Le chiffre pour la coopération totale (4.997 millions € pour 2008-2011) inclut la coopération gouvernementale, la coopération non gouvernementale, la coopération multilatérale, le Fonds belge pour la sécurité alimentaire, l'aide humanitaire et alimentaire, l'aide au secteur privé (BIO) et d'autres frais divers.

Conclusion 2: Potentiel de l'évaluation peu utilisé par la DGD

La décision prise par la DGD de confier presque complètement l'évaluation aux ANG l'a largement privée d'un outil clé pour rendre compte de l'utilisation des fonds publics qui leur étaient confiés, de capitaliser sur les leçons identifiées et de disposer de l'information générée pour alimenter sa prise de décision.

Basée sur les questions Q1 et Q7

Cette conclusion est basée sur les constats suivants:

- La DGD n'avait pas défini de politique d'évaluation générale clarifiant ses objectifs en la matière (cf. Conclusion 1), comme par exemple la reddition de compte, l'aide à la prise de décision, ou la capitalisation sur les connaissances et le partage de leçons. Le cadre réglementaire ne prévoyait pas non plus d'implication (ou pratiquement pas) de la DGD dans le suivi des évaluations. Il ne prévoyait par exemple pas la transmission des rapports d'évaluation par les ANG à la DGD (à l'exception d'une « synthèse de max. 5 pages » pour les « autres partenaires »). (Q1) L'utilisation des rapports d'évaluation dans le dialogue politique avec les ANG n'était pas ou peu formalisée. Il n'y avait pas de politique interne de capitalisation sur base des leçons des évaluations, ni de système pour classer les évaluations (permettant ainsi de les trouver aisément). Peu de capitalisation a ainsi eu lieu.
- La DGD disposait par ailleurs d'une faible capacité en évaluation, due à l'expérience et la formation limitées en évaluation de la plupart des gestionnaires de programmes, ainsi qu'à la disparition d'une fonction interne d'évaluation à la DGD depuis 2010. (Q7) La DGD n'a en fait plus de responsabilité directe en évaluation depuis lors, celle-ci se situant à la CTB (pour la coopération bilatérale), chez les ANG (pour la coopération indirecte) et auprès d'organisations internationales (pour la coopération multilatérale), ainsi qu'au SES, un service externe rapportant directement au Ministre de la Coopération au Développement (pour des évaluations de niveau plus stratégique).
- Une grande majorité des gestionnaires de la DGD considéraient toutefois que les rapports d'évaluation avaient une valeur ajoutée dans leur travail de suivi des programmes des ANG (cf. Conclusion 9), en plus du rapportage narratif et financier par les ANG eux-mêmes. Les gestionnaires les ont ainsi souvent utilisés en pratique dans le dialogue politique avec les ONG et en particulier pour l'appréciation de nouvelles phases de programmes, même si de façon très variable et nettement moins avec les autres acteurs. (Q7)

Conclusion 3: Evaluation utilisée par les ANG pour la prise de décision

Les ANG ont conduit des évaluations principalement dans l'optique d'améliorer la suite de leur programme. Ils l'ont nettement moins utilisée dans une perspective de rendre compte ou de partager les leçons avec de tierces parties.

Basée sur les questions Q1 et Q6

Cette conclusion est basée sur les constats suivants:

- Le cadre réglementaire peu directif de la DGD a permis à la plupart des ANG, en particulier les ONG, les universités et les institutions scientifiques, d'évaluer leur programme de la façon qui leur semblait la plus utile. (Q1)
- Une grande majorité des ANG a utilisé l'évaluation de leur programme comme un outil d'aide à la décision, pour disposer d'une base solide pour revoir l'approche du programme en cours ou pour préparer le suivant. Pas plus de la moitié des ANG n'ont utilisé l'évaluation en pratique pour rendre compte. Par ailleurs, les leçons n'étaient souvent pas partagées au-delà des personnes impliquées directement dans chaque programme. (Q6)

Conclusion 4: Faible capacité à rendre compte

Les évaluations n'avaient qu'une faible capacité à rendre compte. En effet, elles n'ont souvent pas été conçues ni utilisées dans cette optique. Des problèmes de qualité et des questions sur l'indépendance affectaient par ailleurs couramment leur crédibilité. Le cadre réglementaire ne prévoyait pas non plus que les ANG dussent les transmettre à la DGD.

Basée sur les questions Q1, Q4, Q5, Q6 et Q7

Cette conclusion est basée sur les constats suivants:

- Le cadre réglementaire de la DGD n'assignait pas d'objectif particulier à l'évaluation des programmes des ONG, des universités et des institutions scientifiques. Celui pour les « autres partenaires » précisait toutefois un objectif, d'aide à la prise de décision, en indiquant que les évaluations dient être « engagées en vue d'une amélioration prospective », le rôle d'« indiquer les résultats obtenus » étant lui dévolu au rapportage annuel par les ANG eux-mêmes. (Q1)
- Si la plupart des ANG considéraient que reddition de compte est un des objectifs prioritaires de l'évaluation, seule une moitié d'entre eux a intégré cet aspect dans l'évaluation examinée. (Q6)
- Le cadre réglementaire ne prévoyait pas que les ANG dussent transmettre le rapport d'évaluation à la DGD. (Q1) La DGD ne disposait ainsi pas d'une part significative des rapports d'évaluation, en particulier ceux d'ONG. (Q7)
- Analyser les résultats atteints par les différents programmes sur base des rapports existants en vue de rendre compte sur l'ensemble est rendu très compliqué par la grande hétérogénéité des rapports, qui ne permet pas de consolider l'information ; cette diversité s'explique par le peu de directives données par la DGD aux ANG pour l'évaluation. (Q1, Q5, Q7)

- La crédibilité des évaluations, souvent mise en doute, est également un réel problème dans une perspective de reddition de compte. Les éléments affectant la crédibilité des rapports étaient en particulier des problèmes de qualité et/ou des questions sur leur indépendance. (Q4)

Conclusion 5: Peu de partage des leçons des évaluations

Les leçons des évaluations ont rarement été partagées au-delà du cadre de chaque programme.

Basée principalement sur la Q6 ainsi que sur la Q1 et la Q5

Cette conclusion est basée sur les constats suivants:

- La plupart des interlocuteurs ont affirmé que l'« apprentissage » était un objectif prioritaire pour l'évaluation. La plupart d'entre eux entendaient toutefois par là le fait de tirer des leçons sur le programme/projet évalué pour améliorer la *prise de décision* liée à la continuation de ce programme. Peu d'entre eux entendaient par contre par là le fait de *capitaliser sur les connaissances et de partager les leçons* avec un public plus large que celui du programme. (Q1 et Q6, ainsi que l'encadré 2).
- La plupart des ANG considéraient que capitaliser sur les connaissances est un des objectifs prioritaires pour l'évaluation. Ceci n'a en fait vraiment eu lieu que dans une bonne moitié des évaluations examinées. Les leçons d'une évaluation ont régulièrement été partagées en interne et auprès des parties prenantes directes, mais pas systématiquement, et rarement au-delà. L'apprentissage se limitait ainsi typiquement aux acteurs du programme. (Q6)
- Tirer des leçons de l'ensemble des évaluations est rendu compliqué par le fait qu'il n'y avait pas de directive sur la mise en évidence de leçons *transférables*. Il n'était pas aisé d'en identifier et les leçons tirées n'étaient pas toujours utiles en dehors du contexte du projet. La grande hétérogénéité des évaluations et des rapports ne permettait pas non plus une consolidation et une synthèse aisée des leçons apprises. (Q1, Q5, Q6)

L'initiative de la DGD de financer un projet de partage des rapports d'évaluation entre ONG au travers d'une base de données n'a pas eu le succès escompté. De nombreuses ONG sont en fait restées réticentes à partager leurs rapports d'évaluation et à communiquer ouvertement sur les points négatifs. Ceci était nettement moins le cas des universités et institutions scientifiques, qui trouvaient naturelle la transparence sur les résultats vu qu'ils dépendaient largement de financement public, dans une perspective de long terme. (Q6)

Conclusion 6: Progrès modeste des capacités des ANG en évaluation

Les capacités en évaluation d'une majorité d'ANG se sont améliorées au cours de la période 2008-2011. Peu d'entre eux sont toutefois arrivés jusqu'au développement d'un politique interne d'évaluation et de guidance opérationnelle en la matière.

Basée principalement sur la question Q1, ainsi que sur la Q2 et la Q5

Cette conclusion est basée sur les constats suivants:

- Une majorité d'ANG a indiqué que leurs capacités et pratiques en évaluation se sont améliorées par rapport à la période antérieure. La moitié des ONG ont mentionné que ceci était entre autres grâce à l'incitation de la DGD. (Q1)
- L'analyse a toutefois montré que les ANG surestimaient souvent leur capacité en évaluation, ainsi que la qualité des rapports. Près d'un tiers des ANG a ceci dit fait état explicitement d'une augmentation limitée. Les trois quarts ne disposaient par ailleurs pas d'une politique d'évaluation interne bien définie accompagnée d'une guidance opérationnelle pour la mise en œuvre. Les ANG manquaient aussi souvent de références pour apprécier leur politique d'évaluation ou pour pouvoir en définir une. (Q1) Des systèmes performants de suivi et d'évaluation ont en outre peu souvent été mis en place. (Q2) L'encadrement des évaluations a par ailleurs été faible pour la moitié des processus examinés, en particulier pour les ONG. (Q2) La qualité des rapports posait également problème dans une moitié des cas (voir la Conclusion 7 ci-dessous).

Conclusion 7: Qualité faible de nombreux rapports

La qualité des rapports d'évaluation pose problème. Même si elle est bonne dans l'ensemble pour une grande majorité des programmes d'universités et d'institutions scientifiques, elle est faible pour la moitié des programmes d'ONG et d'autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité et s'expliquent par une multitude de facteurs. Ces constats sur la qualité suggèrent que la culture et la maîtrise de l'évaluation restent un défi pour les ANG ainsi que pour nombre d'évaluateurs.

Basée principalement sur la Q2 ainsi que sur les Q1 et Q3

Cette conclusion est basée sur les constats suivants:

- L'appréciation générale de la qualité des rapports d'évaluation est bonne pour la grande majorité des évaluations des programmes des universités et institutions scientifiques (83%). Celle-ci n'était cependant globalement bonne que pour 47% des ONG et 55% des « autres partenaires ». Dans l'ensemble, aucun des rapports examinés n'a été apprécié comme étant de 'très bonne qualité'. (Q3)
- Les faiblesses observées étaient liées à presque tous les critères de qualité. Les problèmes les plus fréquents étaient l'absence d'une revue critique de la conception du programme, des questions d'évaluation peu précises, l'absence de données de suivi fiables, le peu de précisions données sur le degré de fiabilité des rapports et le manque de croisement des données. Autour de 40% des rapports d'ONG et d'autres partenaires étaient par ailleurs faibles en termes de solidité de l'analyse et de validité et utilité des conclusions et recommandations. (Q3)
- Les facteurs expliquant les différences de qualité sont multiples et variés. Citons en particulier les suivants (Q2) :
 - *Demande* : budgets peu réalistes, peu de référence à des critères d'évaluation importants, etc. ;

- *Fiabilité des données*: peu de systèmes de suivi et d'évaluation (M&E) au niveau des actions ; analyses et informations pour lesquelles les évaluateurs ne maîtrisaient pas la qualité⁷⁸ ;
 - *Qualité des évaluateurs* : identification des experts idoines, indépendance d'esprit (vision critique et novatrice), esprit analytique et expertise en rédaction de rapports ; et
 - *Qualité du processus* : en termes d'encadrement des évaluateurs, p.ex. de suivi par un comité de pilotage.
- Ces constats sur la qualité (Q2) et sur les progrès modestes des capacités en évaluation des ANG (Q1) suggèrent que la culture et la maîtrise de l'évaluation reste un défi pour de nombreux ANG.

Conclusion 8: Problèmes dans la conception et la mise en place des programmes

De manière générale, les programmes des ANG ont été confrontés à des problèmes dans la conception de la stratégie d'intervention et dans la mise en place d'un système performant de mesure des résultats et impacts

Basée sur les questions Q2 et Q5

Cette conclusion est basée sur les constats suivants:

- Un problème récurrent soulevé dans les conclusions des évaluations était lié à la conception (*design*) des stratégies. (Q5)
- Des systèmes performants pour planifier et mesurer les résultats et impacts à un niveau détaillé ont peu souvent été mis en place.⁷⁹ Les évaluations n'ont ainsi pas pu disposer de données fiables provenant d'un tel système (Q2 et Q5).

Conclusion 9: Peu de jugement sur la contribution au développement

Les rapports d'évaluation ne permettent pas d'avoir un jugement global sur la contribution au développement par la coopération non gouvernementale.

Basée sur la Q1 et la Q5

Cette conclusion est basée sur les constats suivants:

- La DGD n'avait pas défini d'attentes spécifiques en termes de jugement à rendre par l'évaluation sur la contribution des programmes au développement des populations

⁷⁸ Certaines évaluations consistaient par exemple essentiellement en des synthèses d'auto-évaluations ou de rapports par pays préparés en interne, dans lesquels les évaluateurs n'avaient pas été impliqués (p.ex. en termes d'approche méthodologique, de supervision ou de contrôle de qualité).

⁷⁹ Un système de suivi et d'évaluation a souvent été mis en place par les ANG spécifiquement pour leur programme cofinancé par la DGD dans son ensemble. Il n'y avait toutefois souvent pas de système interne suffisamment détaillé que pour permettre d'évaluer l'atteinte des résultats spécifiques attendus au niveau des projets/activités.

cibles (voir Conclusion 1). Elle n'avait pas non plus défini de stratégie et de dispositif pour permettre la consolidation des résultats des évaluations de tous les programmes, permettant un jugement d'ensemble sur le sujet. (Q1)

- De nombreux rapports d'évaluation examinaient l'efficacité des interventions en termes de bonne conduite des activités et d'atteinte des '*outputs*', sans toutefois examiner les résultats au niveau des destinataires ou des pays partenaires. (Q5)
- Les ANG n'ont pas demandé aux évaluateurs d'examiner la question de la durabilité de leur action dans la moitié des cas, tout comme celle de l'impact. (Q2) Il s'agit pourtant de deux critères d'évaluation essentiels. S'il est souvent difficile de mesurer la durabilité par faute de recul dans le temps, et l'impact par faute de moyens ou de méthode adéquate, il est néanmoins étonnant de constater l'absence de questionnement formel sur la durabilité et l'impact *potentiels* de l'intervention et sur les facteurs les favorisant. Ces deux éléments sont ainsi également peu souvent traités dans les conclusions et les recommandations des rapports. (Q5)
- Les évaluations ne permettent donc pas de donner à un niveau consolidé des réponses à la question de la contribution de la coopération non gouvernementale de la DGD au développement. Ceci est toutefois une question centrale de la réflexion des dernières années sur la coopération au développement, à la base du Partenariat de Busan pour une coopération efficace au service du développement (2011) entre les multiples acteurs de la coopération au développement, dont les bailleurs bilatéraux et les acteurs de la société civile.

Conclusion 10: Évaluations utiles pour les ANG et gestionnaires DGD

Les évaluations ont été considérées comme utiles à la prise de décision par la plupart des ANG. Les gestionnaires de la DGD les ont également largement appréciées, en particulier pour la valeur ajoutée qu'elles offraient par rapport au rapportage annuel par les ANG.

Basée sur les questions Q6 et Q7

Cette conclusion est basée sur les constats suivants:

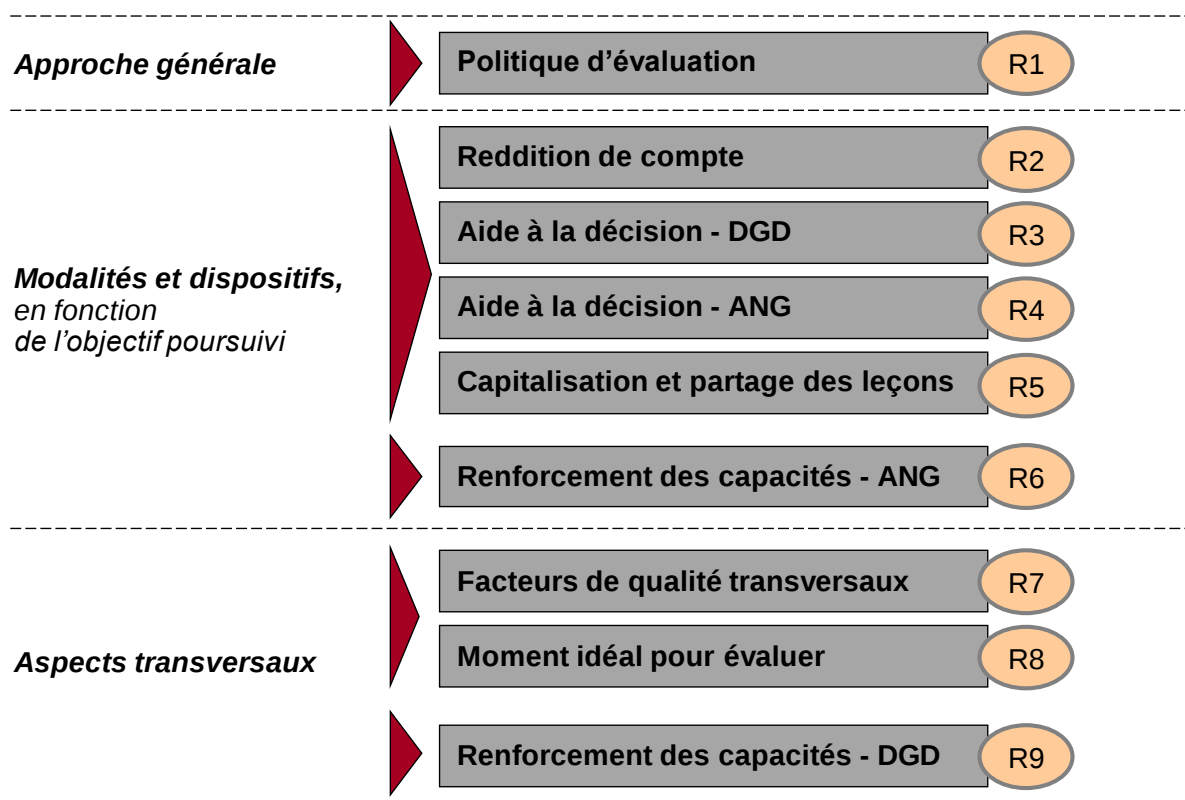
- 84% des ANG ont répondu à l'enquête considérer les recommandations des évaluations comme (très) valides et utiles. Ceci indique une grande satisfaction sur l'utilité de l'évaluation : la plupart des ANG avaient en effet commandité l'évaluation en vue d'améliorer leur programme (suivant), accordant ainsi une grande importance dans leur appréciation à la validité et à l'utilité des recommandations. (Q6)
- 82% des gestionnaires de la DGD ayant répondu à l'enquête ont indiqué que les rapports d'évaluation avaient une valeur ajoutée (35%) voire une grande valeur ajoutée (47%) dans leur travail de suivi des programmes des ANG, en plus du rapportage narratif et financier. Les évaluations avaient selon eux pour principal avantage de fournir une information complémentaire, (plus) indépendante et potentiellement avec un autre point de vue. Certains ont également indiqué que l'évaluation permettait de voir au-delà de la gestion du programme, avec un jugement plus tranché sur l'atteinte des objectifs. (Q7)

5. Recommandations

Ce chapitre présente les recommandations qui émergent de cette méta-évaluation. Comme indiqué dans le cahier des charges, elles visent à « contribuer à la mise en place d’une réglementation cohérente et d’une politique coordonnée en matière de suivi et d’évaluation au sein de la DGD dans le cadre du travail avec les acteurs non gouvernementaux ».

La figure ci-dessous fournit un aperçu des thèmes du set de recommandations. Chacune d’elles est décrite en détail par la suite.

Figure 11 - Recommandations



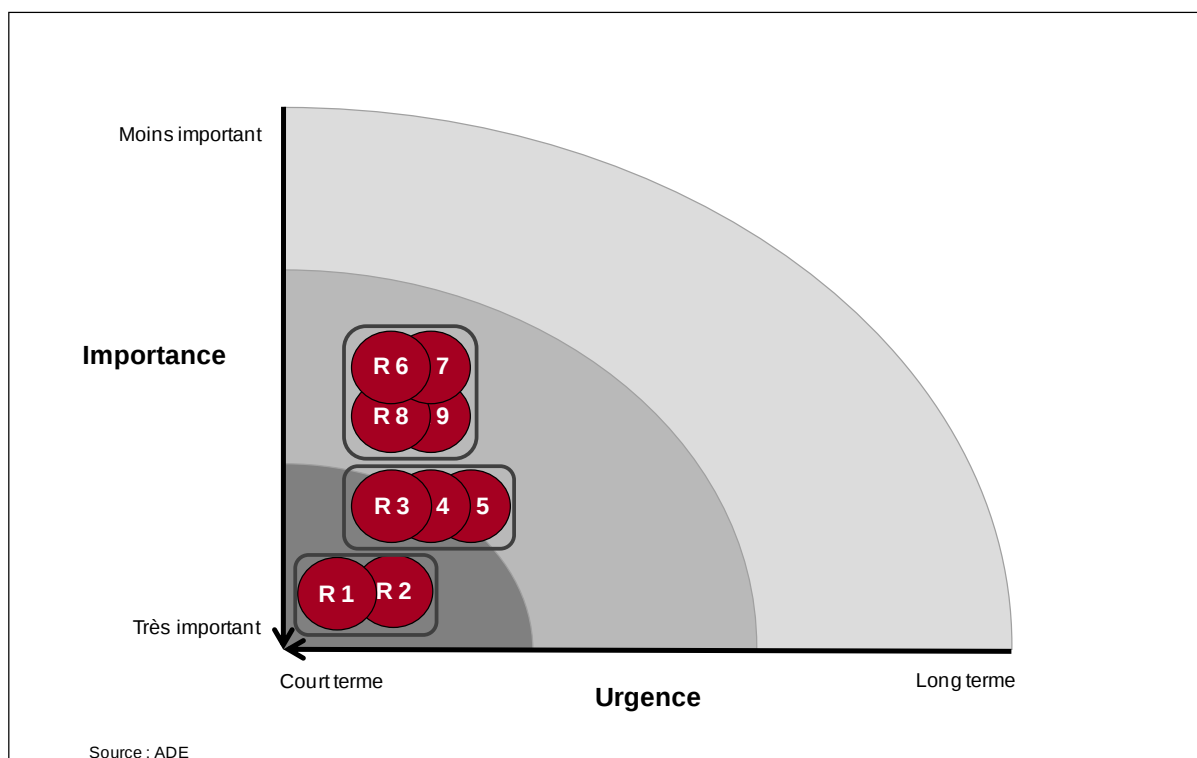
Source : ADE

Chaque recommandation réfère aux conclusions sur lesquelles elle se base. Elles précisent également à qui incombe la responsabilité de leur mise en œuvre.

La figure ci-dessous fournit le niveau de **priorité** de chaque recommandation, en termes d'urgence et d'importance. La plupart des recommandations sont en fait considérées comme importantes et urgentes, dans la mesure où la coopération gouvernementale belge est en train de redéfinir le cadre législatif et réglementaire de ses subventions aux acteurs non gouvernementaux pour la décennie à venir. La nouvelle loi sur la coopération au développement qui vient d'être publiée exige par ailleurs qu'une approche soit élaborée pour l'évaluation des résultats dans une perspective de reddition de compte.

Les recommandations les plus importantes et les plus urgentes sont ainsi de revoir et de clarifier l'approche générale de la DGD pour l'évaluation de sa coopération non gouvernementale, et de développer un système structuré pour rendre compte. (R1 et R2) Les recommandations R3 à R5 sont importantes également pour pouvoir utiliser l'évaluation comme outil d'appui à la décision et de partage de leçons. Les recommandations R6 à R8 fournissent des conseils sur des facteurs spécifiques à prendre en compte pour assurer la qualité et l'utilité des évaluations, de manière transversale. La recommandation sur le renforcement de la capacité de la DGD en évaluation est importante également, et non moins urgente (R9). Cet ensemble de recommandations donne ainsi des pistes essentielles pour définir la politique, les modalités et les dispositions du cadre réglementaire en matière d'évaluation de la coopération non gouvernementale belge ; il inclut de nombreux conseils directement utiles aux ANG également.

Figure 12 - Priorité des recommandations



Recommandation 1 : Définir l'approche générale de la DGD en matière d'évaluation

La DGD doit revoir sa politique d'évaluation de la coopération non gouvernementale pour une reddition de compte crédible et pour améliorer l'efficacité de son aide. Elle doit pour cela élaborer une approche générale claire, définissant les objectifs poursuivis, les modalités, les dispositifs et un cadre réglementaire adéquats, ainsi qu'une allocation appropriée des responsabilités et des ressources.

Adressée à la DGD, en collaboration avec les ANG⁸⁰

Basée sur la C1 et l'ensemble des conclusions

La DGD est actuellement confrontée à des défis importants en tant que bailleur de fonds gouvernemental, et ce d'autant plus que le pays est confronté à une crise budgétaire majeure. Elle a l'obligation de rendre compte de l'utilisation efficace et efficiente de l'argent public qu'elle consacre à la coopération au développement. Ceci comprend son appui à la coopération non gouvernementale, qui représente une part substantielle de son aide (18% ou 901 millions € sur la période 2008-2011). La toute nouvelle loi en la matière précise par ailleurs qu'elle doit élaborer une approche pour l'évaluation des résultats dans une perspective de reddition de compte (voir R2). La DGD doit également améliorer l'efficacité de son aide dans l'esprit des engagements internationaux des dernières années, comme le Partenariat de Busan pour une coopération efficace au service du développement (2011).

La DGD doit ainsi pouvoir répondre à quatre questions fondamentales relatives à son appui à la coopération non gouvernementale :

- *Quel système définir pour rendre compte de sa contribution au développement ?*
- *Sur quelle base établir ses décisions d'appréciation des programmes dans le cadre du nouveau cycle de coopération et de continuation de cet appui au terme de ce cycle ?*
- *Quelle information lui est nécessaire pour sa prise de décision stratégique, et comment l'obtenir ?*
- *Comment s'assurer que l'efficacité de cet appui soit améliorée ?*

L'évaluation est un outil qui a un rôle clé à jouer pour répondre à chacune de ces questions. Elle est en particulier essentielle pour une *reddition de compte* plus crédible et à un niveau plus large sur la contribution réelle de l'aide à un développement durable des populations cibles, au-delà d'une mesure de la bonne mise en œuvre des activités et des 'outputs'. L'évaluation est par ailleurs un outil utile pour disposer d'informations solides nécessaires à la *prise de décision* par la DGD, par les ANG ou par d'autres. Elle peut enfin permettre d'identifier des *leçons* sur la coopération évaluée et former une base solide pour les partager, dans une perspective d'organisation(s) apprenante(s), dans le but de faire bénéficier de nombreux autres programmes de ces enseignements.

La présente étude a montré qu'il y avait une grande confusion sur les objectifs assignés à l'évaluation, et que la DGD n'a pas utilisé l'évaluation de façon réellement pertinente,

⁸⁰ Les références aux ANG dans ce chapitre incluent souvent les fédérations et coupoles d'ONG comme destinataires des recommandations. Notons par ailleurs que la dénomination d'« acteur non gouvernemental » (ANG) utilisée durant la période d'évaluation 2008-2011 a été remplacée depuis lors par celle d'« acteur de la coopération au développement non gouvernementale » (ACNG). A des fins de clarté, nous utilisons de façon uniforme dans ce rapport la dénomination qui était d'application durant la période d'évaluation, soit celle d'ANG.

efficace ni efficiente durant la période 2008-2011. Il est ainsi essentiel que la DGD définisse clairement le(s) rôle(s) qu'elle veut assigner à l'évaluation dans le futur, pour sa propre utilisation et/ou pour les ANG. Il serait ainsi utile qu'elle définisse une politique d'évaluation générale. Celle-ci devrait déterminer quel type d'évaluation est amené à répondre à quel type de besoin ; par exemple en termes de reddition de compte (sur quoi ? vis-à-vis de qui ? quel lien avec le système de rapportage général ?), d'aide à la décision (pour qui ? sur quoi ? à quel moment?), ou de capitalisation et de partage des leçons (avec qui ? sur quoi ? comment ?). La politique d'évaluation devrait par ailleurs définir les modalités, les dispositifs et un cadre réglementaire adéquats, ainsi qu'une allocation appropriée des responsabilités et des ressources, pour mener ceci à bonne fin. Il est pour cela essentiel que la DGD définisse la part de responsabilité qui lui incombe en évaluation, ainsi que celle que les ANG pourraient prendre.

Plusieurs autres bailleurs (ainsi que certains ANG) ont entamé récemment un tel processus de réflexion et de formalisation de leur (nouvelle) politique d'évaluation en matière de développement. Ceci est par exemple le cas de la Commission européenne, de la France ou du Royaume-Uni⁸¹. C'est en lien avec le premier principe de base en matière d'évaluation du Comité d'aide au développement de l'OCDE, qui stipule que les organismes d'aide doivent disposer d'une politique d'évaluation claire.⁸² Pour ce faire, la DGD peut s'appuyer sur ses propres capacités en gestion d'évaluation, ou, mieux, sur une collaboration avec des experts en évaluation (voir à ce propos la R9 ci-dessous). Une concertation avec les ANG également serait certainement utile.

La continuation de la situation actuelle en Belgique caractérisée par l'absence d'une véritable politique d'évaluation ne semble pas une option, ne fût-ce que par les dispositions de la nouvelle loi exigeant le développement d'une approche d'évaluation permettant de rendre compte sur les résultats. Par ailleurs, cela amènerait probablement le même genre de problèmes que ceux observés durant la période d'évaluation, tels qu'une faible aptitude des évaluations à la reddition de compte et à la prise de décision sur base d'information de qualité, et un partage limité des leçons pour améliorer l'efficacité et l'efficience de l'aide.

Les recommandations ci-après peuvent servir de pistes pour concevoir les modalités d'une politique d'évaluation et pour définir les dispositions du cadre réglementaire, sur base des enseignements de cette étude. Celles-ci ont trait en particulier à la reddition de compte (R2), à l'utilisation de l'évaluation pour l'aide à la décision par la DGD (R3) et par les ANG (R4), et à l'utilisation de l'évaluation pour capitaliser et partager les connaissances (R5). Les recommandations R6 et R9 ont trait au renforcement nécessaire des capacités tant respectivement des ANG que de la DGD. Les recommandations R7 et R8 fournissent des facteurs spécifiques de qualité et d'utilité à prendre en compte également lors de la définition de la politique d'évaluation.

⁸¹ UK Department for International Development, May 2013. *International Development Evaluation Policy*.

⁸² « Les organismes d'aide devraient avoir une politique d'évaluation s'accompagnant de lignes directrices et de méthodes clairement énoncées, et ils devraient préciser le rôle et les tâches de l'évaluation ainsi que la place de cette fonction dans les structures institutionnelles. » (OCDE, Le Réseau du CAD sur l'évaluation en matière de développement, 2011. *Récapitulatif des normes et standards de référence, Deuxième édition*, p7).

Recommandation 2 : Développer un dispositif structuré pour rendre compte

La DGD devrait créer un dispositif structuré pour permettre une reddition de compte plus crédible et plus utile par l'évaluation.

Adressée à la DGD, en collaboration avec les ANG

Basée sur les conclusions C1, C2 et C4

La DGD a l'obligation en tant que bailleur de fonds gouvernemental de rendre compte au Parlement et aux contribuables de l'utilisation de l'argent public qu'elle consacre à sa coopération, y compris via des acteurs non gouvernementaux. Elle est aussi tenue de rendre compte sur les résultats aux destinataires des actions de coopération et aux autres parties prenantes, conformément à ses engagements internationaux.⁸³ La crédibilité de cette reddition de compte est importante, de façon générale et en particulier dans le cadre des enjeux actuels, dans un contexte de coupes budgétaires et de mise en doute de la contribution de l'aide à un développement réel des pays partenaires. L'évaluation externe n'a pas été utilisée pour la reddition de compte dans le cadre de la coopération non gouvernementale belge. La DGD rendait compte sur base des informations fournies directement par les ANG, au travers de leur rapportage narratif et financier, avec les limitations que cela implique en termes de crédibilité et de niveau stratégique (cf. infra). Il ressort en outre de cette méta-évaluation que les évaluations réalisées n'avaient souvent qu'une faible capacité pour rendre compte de façon crédible (voir C4). Par ailleurs, une évaluation récente du rapportage des résultats a montré des lacunes importantes en matière de rapportage (sur les résultats) de la DGD, dues par exemple à l'absence d'un système holistique de rapportage à un niveau stratégique supérieur. Une grande majorité des gestionnaires de la DGD considérait néanmoins que les évaluations externes qui leur avaient été transmises avaient une valeur ajoutée par rapport au rapportage narratif et financier par les ANG, fournissant une information complémentaire, (plus) indépendante et allant au-delà de la gestion du programme, avec un jugement plus tranché sur l'atteinte des objectifs et sur les réflexions.

L'évaluation externe est en effet un outil important pour une reddition de compte crédible sur les résultats, à nos yeux, ainsi qu'à ceux d'un grand nombre de bailleurs gouvernementaux (voir R1) et d'autres acteurs. L'intention est affichée depuis peu d'intégrer (partiellement) l'évaluation dans le système de reddition de compte sur la coopération belge. La toute nouvelle Loi relative à la Coopération au Développement (datée du 19 mars 2013 et publiée au Moniteur belge du 12 avril 2013)⁸⁴ précise en effet qu'une approche cohérente doit être élaborée pour l'évaluation des résultats, « en vue de permettre un rapportage des résultats et une gestion axée sur les résultats ».⁸⁵ La

⁸³ Voir par exemple la déclaration du 4^e Forum à haut niveau sur l'efficacité de l'aide, 2011. *Partenariat de Busan pour une coopération efficace au service du développement*, 2011, p. 3 (cf. bibliographie en Annexe 14).

⁸⁴ Un Projet de loi modifiant la loi du 19 mars 2013 relative à la Coopération au Développement a été adopté le 17 juillet 2013 en séance plénière de la Chambre des représentants de Belgique et a été transmis au Sénat. Les modifications proposées ne changent toutefois pas sur le fond les dispositions en matière d'évaluation des résultats.

⁸⁵ « CHAPITRE 9. — Évaluation. 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 la 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

DGD a également répondu en ce sens dans sa réponse managériale à la récente évaluation de son rapportage sur les résultats, précisant que le système d'évaluation des interventions « doit permettre un rapportage aussi bien spécifique qu'agrégé des résultats ».⁸⁶ Elle y indique toutefois vouloir donner la responsabilité de la gestion de la mise en œuvre de ce système aux partenaires exécutants (ce qui inclut logiquement les acteurs non gouvernementaux).

Sur base des conclusions de la présente méta-évaluation, il est recommandé à ce propos de considérer le set (indivisible) d'aspects suivants, en précisant que la question de l'utilisation de l'évaluation comme outil pour rendre compte devra être étroitement lié et intégré à la réflexion d'ensemble et au dispositif général à définir par la DGD pour une gestion et un rapportage axés sur les résultats :

- **R2.1 – 'Représentativité'** : Le système à définir doit permettre de rendre compte de l'atteinte des résultats en couvrant une proportion suffisamment large de la coopération non gouvernementale de façon à être crédible sur l'ensemble. Plusieurs systèmes sont envisageables. La DGD pourrait requérir qu'une part minimale des interventions de chaque ANG soit évaluée, comme le demandent les Pays-Bas par exemple. Cela pourrait également se faire en déterminant un protocole basé sur un échantillon d'ANG dont les projets seraient évalués de façon indépendante dans une optique de reddition de compte. Un tel échantillon pourrait être déterminé de façon aléatoire, sur base de critères, ou sur base volontaire. Il est également envisageable d'organiser la reddition de compte sur base d'une série d'évaluations géographiques ou thématiques, qui couvriraient chacune l'action de plusieurs ANG, et ensemble une part significative de la coopération non gouvernementale belge.
- **R2.2 – Évaluation ex post** : La reddition de compte sur les résultats devrait typiquement se faire après la clôture des activités des interventions, de façon « ex post ». Ceci permettrait de disposer d'une base plus solide et fiable pour rendre compte des résultats obtenus. Dans la nouvelle optique de programmation sur 10 ans, consistant en deux programmes successifs d'une durée de 5 ans, les évaluations visant la reddition de compte auraient donc lieu en tout cas après la période de 10 ans et éventuellement après les premiers programmes de 5 ans. Ceci pourrait inclure des évaluations à conduire plusieurs années après la fin des activités, dans l'optique de rendre compte de la contribution au développement dans la durée.
- **R2.3 – Crédibilité** : La crédibilité des évaluations visant la reddition de compte devra être une préoccupation importante. Il faudra en particulier faire attention à la **qualité** et à la pertinence des évaluations, tant en termes de rapports que de processus (voir la R7). Il faudra également veiller à la perception d'**indépendance** des évaluations par le public visé (Parlement, contribuables, destinataires, etc.) : indépendance des commanditaires, des prestataires, du processus, etc. La DGD devrait ainsi avoir un rôle actif dans les évaluations visant la reddition de compte, que ce soit en tant que commanditaire, via une structure externe (comme par exemple du type du dispositif décrit dans la R6), par son implication dans l'encadrement du processus d'évaluation (p.ex. par sa participation à un comité de pilotage), ou d'une autre façon. Enfin, il faudra considérer la **transparence** des évaluations, en termes de processus, d'implication des différentes parties, de publication des rapports (individuels ou synthétiques), etc.

futur des actions entreprises. » (État belge, 2013. *Loi relative à la Coopération au Développement*, p. 22569).

⁸⁶ « Au sein d'une DGD réformée, un projet sera lancé sur recommandation du service Qualité et Résultats et avec le consentement du comité stratégique, afin que soit développé un système d'évaluation des interventions. Dans ce cadre, les efforts viseront, dans la mesure du possible, à privilégier l'uniformisation de la méthode, sans toutefois perdre de vue la spécificité des différents canaux d'exécution. Ce système doit permettre un rapportage aussi bien spécifique qu'agrégé des résultats. La mise en place de ce système se déroulera en concertation avec les partenaires exécutants, qui gèreront en outre la mise en œuvre de ce système. » (Réponse managériale de la DGD à l'évaluation du système de rapportage des résultats de la DGD, dans Service de l'Évaluation spéciale, 2012, p. 118).

- **R2.4 – Consolidation:** La DGD devra déterminer clairement ses besoins pour répondre aux attentes de reddition de compte du Parlement et des contribuables, ainsi que des destinataires de l'aide et des autres parties prenantes. Il est important d'assurer que le système permette de synthétiser et de consolider de façon la plus aisée possible les informations provenant des diverses évaluations visant la reddition de compte. Ceci afin qu'elles puissent fournir une vision d'ensemble des résultats obtenus. Cela rejoint l'intention de la DGD exprimée dans sa réponse managériale à la récente évaluation de son rapportage sur les résultats (cf. supra). Un certain degré d'harmonisation sera indispensable. Il est néanmoins recommandé de baser dans la mesure du possible le système de reddition de compte sur les politiques et systèmes de suivi et d'évaluation des ANG, ce qui implique également de la flexibilité. Une concertation avec les acteurs non gouvernementaux est nécessaire à cette fin.
- **R2.5 – Expertise en évaluation:** Il est important que la politique et le système d'évaluation de la DGD soient les plus pertinents, efficaces et efficients possibles, et qu'ils aient la crédibilité nécessaire dans une optique de reddition de compte. Il est ainsi recommandé à la DGD de faire appel à de l'expertise pointue en la matière (comme détaillé dans la R9). Par ailleurs, dans le cas où le système d'évaluation envisagé par la DGD prévoirait des évaluations visant la reddition de compte au niveau des interventions des ANG avec une grande responsabilité de ces mêmes acteurs dans leur gestion, il est recommandé à la DGD de définir des dispositions claires et précises sur ses besoins et attentes (en matière de qualité, d'indépendance, de transparence, etc.). Le respect de ces dispositions par les ANG (et leurs évaluateurs) pourrait être vérifié par des experts en évaluation, comme ceux indiqués dans la R9 (internes, SES, externes, etc.). Une meilleure option serait encore de donner cette responsabilité à un dispositif structurel associant en premier lieu la DGD et les ANG (comme celui recommandé dans la R6).

Le cadre réglementaire défini sur base de ce set de recommandations devra être précis et univoque, afin d'éviter les besoins de clarification et les différences d'interprétation observées durant la période 2008-2011. Il devra également s'inscrire autant que possible dans le processus de 'simplification administrative' engagé par la DGD et tenir compte des besoins et attentes des acteurs non gouvernementaux. Il est ainsi recommandé de collaborer étroitement avec ceux-ci lors de la définition de ce cadre réglementaire.

L'absence d'un système d'évaluation axé sur la reddition de compte sur les résultats qui soit adéquat et permette la consolidation, limiterait clairement la capacité de la DGD de rendre compte des résultats de sa coopération non gouvernementale de façon crédible et synthétique, comme ce fut le cas durant la période d'évaluation 2008-2011.

Recommandation 3 : Permettre à la DGD d'utiliser l'évaluation comme outil d'aide à la décision	
La DGD devrait avoir la possibilité d'utiliser l'outil qu'est l'évaluation pour ses propres besoins d'aide à la décision.	
<i>Adressée à la DGD</i>	
<i>Basée sur la conclusion C2</i>	

Une distinction est faite pour les évaluations visant un objectif d'aide à la décision en fonction des décideurs concernés, en particulier le pouvoir politique et la DGD (traités dans la présente R3) et les acteurs non gouvernementaux (traités dans la R4 ci-après).

La responsabilité de l'évaluation des programmes des acteurs non gouvernementaux à été donnée à ces mêmes acteurs durant la période d'évaluation 2008-2011. Le pouvoir politique (Ministre et Parlement) pouvait toutefois disposer d'évaluations à un niveau plus stratégique au travers des études commanditées par le SES, qui est un service externe rapportant directement au Ministre de la Coopération au Développement, avec ses propres budget et programme d'évaluation. La DGD avait également la capacité de commanditer de telles études au travers de son service d'évaluation interne, mais ceci a cessé en 2010 avec la suppression de ce service, qui a été largement intégré au SES. La DGD ne dispose ainsi à l'heure actuelle (pratiquement) plus de la possibilité d'utiliser l'évaluation comme un outil pour ses propres besoins d'aide à la décision stratégique. Ceci malgré le fait qu'un rôle justement plus stratégique est assigné à la DGD au sein de la coopération belge. Il est ainsi recommandé à la DGD d'envisager activement de se doter de la possibilité de pouvoir commanditer à nouveau elle-même des évaluations sur des sujets importants pour sa propre prise de décision. Si cela pourrait se faire par une capacité interne en évaluation (avec un mandat complémentaire à celui du SES), il est également envisageable par exemple de trouver une façon de commanditer des évaluations via ou avec le support du SES, ou via des dispositifs spécialisés en évaluation, gérés par des externes ou conjointement entre la DGD et les ANG, comme via un dispositif du type de celui proposé dans la R6. Concernant l'aide à la décision au niveau des programmes individuels, il est recommandé que les gestionnaires de la DGD puissent s'appuyer sur les évaluations focalisées sur la reddition de compte (avec un rôle certain de la DGD – cf. R2) et sur les évaluations focalisées sur l'aide à la décision par les ANG (avec une interaction limitée de la DGD – cf. R4).

Recommandation 4 : Laisser de la flexibilité pour l'aide à la décision des ANG

Une liberté d'action importante devra être donnée aux acteurs non gouvernementaux pour les évaluations qui visent à les aider dans leur propre prise de décision.

Adressée à la DGD, en collaboration avec les ANG

Basée sur les conclusions C3 et C7

Les acteurs non gouvernementaux (ANG) ont principalement utilisé l'évaluation durant la période 2008-2011 dans une perspective d'aide à leur prise de décision. La plupart d'entre eux utilisaient l'évaluation pour améliorer leur programme ou actions en cours ou le(s) suivant(s). Ils ont largement apprécié l'utilité des évaluations conduites, malgré leurs faiblesses. Il a ainsi beaucoup de sens de continuer à promouvoir l'évaluation comme outil d'aide à la décision pour les ANG, tout en veillant à la qualité. Il est important dans cette perspective que les acteurs puissent utiliser au mieux l'évaluation en fonction de leurs besoins. Il est ainsi recommandé de laisser une grande flexibilité d'action aux acteurs non gouvernementaux, pour en définir le sujet, la méthodologie, l'équipe d'évaluation, etc. Ceci pourrait impliquer l'utilisation ou le développement d'approches innovantes. Si une large flexibilité est conseillée, il faudra néanmoins trouver un moyen, en collaboration avec les acteurs en question, de favoriser la qualité des évaluations (voir la R7 ci-dessous). Le moment et la fréquence de ce type d'évaluation devraient idéalement aussi être laissés largement à l'appréciation des acteurs, en fonction de leurs besoins. Un équilibre devra en cela être trouvé entre les évaluations visant l'aide à la décision pour les ANG et celles (éventuelles) visant la reddition de compte (cf. R2), le cas échéant. L'interaction de la DGD dans ces processus d'évaluation devra par ailleurs être limitée (sauf si désirée par l'acteur en question), afin d'éviter les effets contre-productifs, tels qu'une autocensure sur des aspects sensibles, ou l'inclusion de questionnements répondant plus aux besoins de la DGD que des acteurs. Il est toutefois raisonnable que la DGD puisse disposer des rapports d'évaluation finaux afin qu'elle puisse elle-même être informée des leçons du programme et en tenir

compte (au moment opportun) dans sa propre prise de décision. Il est également envisageable de proposer que les évaluations pour les ANG incluent accessoirement des recommandations à la DGD quand cela s'avère utile, comme ce fut observé dans quelques cas (voir section 3.7.2).

Recommandation 5 : Élaborer un dispositif actif pour le partage des leçons
La politique d'évaluation de la DGD devrait inclure un dispositif visant à favoriser un large partage des leçons.
<i>Adressée à la DGD et aux ANG</i>
<i>Basée sur la conclusion C5</i>

Les leçons décrites dans les différentes évaluations n'étaient pas toujours aisément identifiables, ni transférables en dehors du contexte du programme. Par ailleurs, les évaluations ont en général peu été transmises au-delà des acteurs directs du programme. Les leçons de la coopération non gouvernementale durant la période 2008-2011 n'ont ainsi généralement pas pu bénéficier à un grand nombre d'acteurs.

Capitaliser sur les connaissances et partager les leçons apprises est toutefois l'un des trois grands types d'objectifs possibles de l'évaluation. En tant que bailleur gouvernemental finançant un grand nombre d'interventions, souvent dans les mêmes pays, secteurs ou avec des problématiques similaires, il a beaucoup de sens que la DGD soutienne activement l'accumulation et le partage des connaissances en vue de favoriser l'efficacité de son aide. Ceci est également un engagement de la société civile. Les Principes d'Istanbul sur l'efficacité du développement des organisations de la société civile (2010), promeuvent ainsi explicitement le « partage des connaissances et l'apprentissage mutuel »⁸⁷ ainsi que la « transparence »⁸⁸.

Il est dès lors recommandé de développer, en collaboration avec les ANG, un dispositif pour favoriser activement la capitalisation des connaissances et le partage des leçons. Il conviendra de distinguer deux types d'actions: (i) des initiatives destinées principalement à cette fin, tels que des mécanismes et exercices de capitalisation et de diffusion (évaluations spécifiques, méta-analyses, séminaires, bases de données, etc.) ; et (ii) la définition de dispositions pour favoriser la mise en exergue de leçons transférables au sein des évaluations qui visent principalement la reddition de compte ou l'aide à la décision. Il faudra également favoriser activement une *culture* de partage et d'échange de leçons avec l'extérieur. Malgré les intentions affichées dans les Principes d'Istanbul (2010), de nombreux ANG restent en effet réticents à la diffusion de messages négatifs sur leur action. Ce faisant, il faudra veiller à éviter les aspects contre-productifs d'une transparence non appropriée, en particulier dans le cas d'évaluations commanditées par un acteur pour l'aider dans sa propre prise de décision (une exigence de publication pourrait par exemple favoriser l'autocensure).

⁸⁷ « Les OSC sont efficaces en tant qu'acteurs du développement lorsqu'elles... améliorent la façon de tirer un enseignement de leurs expériences, des autres OSC et acteurs du développement, en intégrant les faits avérés de la pratique et des résultats du développement, et en incluant le savoir et la sagesse des communautés locales et indigènes, afin de renforcer l'innovation et la vision du future qu'elles souhaiteraient construire. »

⁸⁸ « Les OSC sont efficaces en tant qu'acteurs du développement lorsqu'elles... font preuve d'un engagement organisationnel soutenu à la transparence (...) ».

Il serait également utile de prévoir un mécanisme pour favoriser le partage d'expérience entre ANG (voire avec la DGD et le SES) sur l'identification d'évaluateurs compétents.⁸⁹ Ceci pourrait par exemple se faire via une base de données, un réseau ou une plateforme, que ce soit au niveau national ou intégré dans un mécanisme international.

Les modalités spécifiques du dispositif de partage de leçons devront être définies en étroite collaboration avec les ANG, qui en seront les principaux contributeurs et destinataires directs. La mise en œuvre du dispositif décrit dans cette recommandation pourrait se faire par exemple au travers d'un mécanisme conjoint entre l'administration et les acteurs tel que celui envisagé pour le renforcement des capacités des acteurs (voir R6). Il serait en outre certainement utile d'exploiter le potentiel de diffusion au-delà des ANG belges et de viser à en faire bénéficier également le personnel de la DGD, de la CTB et tout public intéressé en Belgique et dans d'autres pays du Nord, ainsi que dans les pays du Sud. Ceci permettrait de faire profiter un maximum d'autres projets, programmes et politiques de l'expérience acquise.

L'absence de mise en place d'un mécanisme effectif promouvant activement l'accumulation et le partage des connaissances avec un public plus large que celui de chaque programme aura probablement pour conséquence d'obtenir (pratiquement) aussi peu de partage de leçons de la coopération non gouvernementale belge que ce qui a été observé durant la période 2008-2011. Ceci serait une occasion manquée en termes d'instrument pour améliorer l'efficacité de l'aide.

Recommandation 6 : Élaborer un dispositif fort pour le renforcement des capacités
La DGD et les acteurs non gouvernementaux doivent s'assurer que l'ensemble des acteurs aient des capacités solides en évaluation, en ce compris en termes de système interne de suivi et d'évaluation performant.
<i>Adressée à la DGD et aux ANG</i>
<i>Basée sur les conclusions C6 et C8</i>

Il y a des indications que la DGD a contribué en partie au renforcement des capacités en évaluation de la moitié des ONG. Ceci suite aux dispositions de son cadre réglementaire, même si peu exigeantes, ainsi que par les formations organisées par les fédérations d'ONG qu'elle finançait entre autres pour renforcer la qualité professionnelle des ONG sur le plan méthodologique. Ceci était toutefois bien moins le cas pour les autres types d'acteurs. Il ressort également de cette étude que le progrès a été modeste. Si les ANG (y compris les ONG) ont largement reçu de la DGD la responsabilité de l'évaluation de leur programme, peu d'entre eux sont arrivés jusqu'au développement d'un politique interne d'évaluation et de guidance opérationnelle en la matière. Des systèmes internes performants de mesure du progrès de l'atteinte des résultats attendus ont par ailleurs peu souvent été mis en place par les ANG à un niveau de qualité et de détail suffisant que pour obtenir des informations fiables pour l'évaluation.

Il est recommandé à la DGD de concevoir, en étroite collaboration avec les acteurs non gouvernementaux, un dispositif *fort* pour s'assurer que l'ensemble des acteurs aient des capacités solides en évaluation. Il est en effet essentiel de renforcer les capacités en évaluation des ANG, en particulier si la DGD continue à leur donner une grande responsabilité en la matière. En l'absence de mesures fortes, le risque est en effet élevé d'arriver au même constat que pour la période d'évaluation 2008-2011, soit un progrès limité des capacités des ANG en évaluation et des faiblesses persistantes dans les systèmes de suivi et d'évaluation. Ceci affecterait à nouveau la pertinence et la qualité

⁸⁹ Cela se fait par exemple en France au travers du F3E (voir R7).

des évaluations, et aussi l'efficacité des budgets y relatifs. Cela devrait se répercuter à son tour sur la qualité des décisions à prendre par les ANG et/ou la DGD, et sur la qualité (et la crédibilité) de la reddition de compte.

Deux options sont envisagées en particulier en termes de dispositif fort, qui seraient idéalement suivies conjointement, en se renforçant mutuellement :

- *Des exigences strictes* sur la capacité des acteurs en évaluation, sur la qualité des rapports d'évaluation et sur le développement d'un système interne de planning, de suivi et d'évaluation performant ;
- *De l'appui au renforcement des capacités des acteurs en évaluation.*

L'établissement d'un tel dispositif devrait se faire en étroite collaboration entre la DGD et les ANG. Le dispositif en question pourrait également consister en un partenariat actif entre ces deux types de parties ; ceci aurait pour avantage de renforcer les capacités des deux côtés, de faciliter l'échange au sujet de l'évaluation, d'utiliser un même langage et de pouvoir répondre directement aux différents types de besoins. Les fédérations et/ou coupoles d'ONG pourraient jouer un rôle dans un tel dispositif. Il faudra veiller à ce que le dispositif permette d'une façon ou d'une autre un renforcement des capacités également des autres types de partenaires (universités, institutions scientifiques, syndicats, pouvoirs locaux, etc.).

Des exemples intéressants existent à ce propos dans d'autres pays, notamment chez nos voisins proches. En **France** par exemple, un dispositif spécifique a été créé en 1994 : le F3E.⁹⁰ Celui-ci émane de la volonté partagée des ONG et des pouvoirs publics français de renforcer les capacités méthodologiques d'évaluation et d'analyse des pratiques des ONG. Il s'est ensuite ouvert à d'autres acteurs: les collectivités territoriales et les établissements publics de santé. Il reçoit un financement substantiel de la coopération française et est basé sur un dialogue paritaire entre l'AFD et les différents acteurs. Le F3E est aujourd'hui un réseau qui vise la professionnalisation des acteurs et les accompagne dans l'amélioration de l'impact et de la qualité de leurs actions. Il joue un rôle de catalyseur et de tiers exigeant. Il anime par ailleurs un espace d'échange et d'apprentissage, développe une fonction de centre de ressources, et promeut l'innovation. Les **Pays-Bas** ont eux chargés en 2003 leur service d'Inspection pour la coopération au développement et la politique d'évaluation (IOB) de veiller à la qualité méthodologique des évaluations des programmes cofinancés. L'approche définie inclut entre autres un seuil minimal de couverture de chaque programme à évaluer et un seuil minimal de qualité défini pour les rapports d'évaluation sur base d'une série de critères. Le respect de ces seuils est vérifié par l'IOB. La coopération néerlandaise affirme que ceci a entraîné un renforcement clair des capacités des ONG en évaluation.⁹¹ Ceci fut confirmé par d'autres parties, qui ont noté que cette approche avait également entraîné un renforcement de capacités parmi les évaluateurs. La politique d'évaluation gouvernementale prévoyait par ailleurs un appui au renforcement des capacités du secteur. La fédération néerlandaise des ONG de coopération internationale avait en outre encadré une initiative de mise en commun de fonds d'un grand nombre d'ONG pour favoriser certaines actions en matière d'évaluation.⁹² Une grande majorité d'entre elles (19 sur 20 'alliances de cofinancement') se sont engagées à faire évaluer conjointement leurs programmes, selon une méthodologie conforme aux exigences de qualité de la coopération néerlandaise.⁹³

⁹⁰ <http://f3e.asso.fr> .

⁹¹ Ministerie van Buitenlandse Zaken, 2012. *Betreft IOB-rapport "De methodische kwaliteit van programma-luities in het Medefinancieringsstelsel-I (2007-2010)"*.

⁹² <https://partos.nl/actueel/nieuws/update-gezamenlijke-mfs-ii-luities-reactie-ministerie-van-buitenlandse-zaken>

⁹³ Ministerie van Buitenlandse Zaken, 2012, p3.

Recommandation 7 : Favoriser les facteurs de qualité
L'évaluation doit tenir compte des facteurs principaux affectant la qualité, tels que la qualité de la demande (objectifs, budget, etc.), des systèmes de suivi et d'évaluation performants au niveau des projets, la qualité des évaluateurs et la qualité du processus.
<i>Adressée à la DGD et aux ANG</i>
<i>Basée sur la conclusion C7</i>

La qualité des rapports d'évaluation posait problème durant la période d'évaluation 2008-2011 pour la moitié des programmes d'ONG et d'autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité et s'expliquent par une multitude de facteurs (voir Conclusion 7).

Il est recommandé à ce propos de prendre des mesures par rapport aux principaux facteurs de qualité identifiés au niveau d'une évaluation individuelle. Il est ainsi conseillé à la DGD de les intégrer dans ses directives, et aux ANG d'en tenir compte directement dans leur politique et pratiques d'évaluation. Ces recommandations sont, sauf avis contraire, valables transversalement pour tous les types d'évaluation :

- **R7.1 – Objectifs** : Définir clairement l'**objectif prioritaire** de toute évaluation, en précisant clairement le type d'objectif (reddition de compte, aide à la décision, ou capitalisation et partage des connaissances) et le public cible (la DGD, l'ANG, le grand public, etc.).⁹⁴ Un second objectif de priorité moindre peut éventuellement être précisé, tout en veillant à circonscrire les effets contreproductifs. Nous déconseillons de définir un troisième objectif ou davantage. Ceci amène à clarifier et optimiser la définition du champ, les questions d'évaluation, la méthodologie, le choix des profils d'évaluateurs, etc.
- **R7.2 – Budget**:
 - Le budget et le calendrier d'une évaluation doivent être **réalistes** pour s'assurer d'une évaluation de qualité.
 - Les **frais relatifs à l'ensemble des activités d'évaluation** d'un programme d'ANG devraient être estimés et pris en compte ; la DGD devrait clarifier ceux qu'elle considère comme éligibles.⁹⁵
 - Une solution structurelle doit être trouvée pour éviter les effets pervers du fait de considérer les frais d'évaluation comme des coûts de gestion du programme, comme ce fut le cas pour les ONG et « autres partenaires » (cf. Encadré 1). Ces frais pourraient par exemple faire l'objet d'une **ligne budgétaire distincte**, comme ce qui était d'application pour les universités et institutions scientifiques.
 - La limitation à un **niveau maximum d'honoraires** par jour devrait être **supprimée** pour éviter les effets contreproductifs (cf. Encadré 1).

⁹⁴ Cf. Encadré 2 dans la section 3.1.1.

⁹⁵ Par exemple : frais d'étude de faisabilité, d'évaluation ex ante, de définition de *base line*, de suivi (monitoring), de formation en évaluation, d'études de visibilité, d'évaluations multi-acteurs, de ressources internes, etc.

- Des **approches efficaces** visant à réduire les frais d'évaluation à charge du programme peuvent s'avérer utiles, comme le recours à du personnel interne, local ou à des étudiants pour effectuer une partie du travail. Comme l'étude l'a montré, il est toutefois important pour la qualité de l'évaluation que l'ensemble soit clairement encadré par les évaluateurs, et qu'ils définissent ou approuvent l'approche méthodologie et les différents outils.
- **R7.3 – Fiabilité des données:** Les évaluations devraient le plus possible pouvoir se baser sur des données fiables relatives à l'atteinte de résultats, notamment des indicateurs au niveau des produits, des réalisations, de l'impact, etc. Ceci proviendrait d'un système de planning, de suivi et d'évaluation de l'ANG au niveau de ses actions – voir la R6 ci-dessus.
- **R7.4 – Méthodologie:** Les considérations méthodologiques pour chaque évaluation devraient dépendre en grande partie de la spécificité de chaque situation, du budget mis à disposition et du type d'objectif poursuivi en priorité (reddition de compte, aide à la décision ou capitalisation). Quelques recommandations transversales clés ressortent néanmoins de l'exercice:
 - **Examiner la logique d'intervention :** la conception du programme n'a pas été examinée dans plus de la moitié des évaluations. Il s'agit pourtant d'un aspect clé de la pertinence de l'action, et de là de son efficacité, de l'efficience et de l'impact attendus. Il est ainsi recommandé d'inclure systématiquement un questionnement critique des liens de cause à effet escomptés par l'action évaluée, avec identification des hypothèses sous-jacentes.
 - **Prendre en compte les critères d'évaluation adéquats:** les critères d'évaluation du Comité d'aide au développement de l'OCDE n'ont pas été utilisés de façon systématique, alors qu'ils sont généralement admis comme les critères de base pour l'évaluation. D'autres critères et principes ont également été peu utilisés, comme ceux relatifs à la Déclaration de Paris ou aux Principes d'Istanbul, ou des facteurs de qualité tels que la satisfaction des bénéficiaires et celle d'autres parties prenantes. Il était par ailleurs rarement demandé d'examiner des thèmes transversaux tels que le genre ou l'environnement. Il n'est en soi pas nécessaire d'inclure chacun de ces 'critères' dans chaque évaluation, au risque de manquer de focus. Il est néanmoins recommandé d'encourager une connaissance plus approfondie de ces 'critères' et de systématiser la réflexion sur le choix des 'critères' les plus pertinents pour chaque évaluation, au moment de la rédaction des termes de référence.
 - **Examiner les 'résultats', la durabilité et l'impact (potentiels) des interventions :** De nombreux rapports d'évaluation ne dépassaient pas la mesure du niveau des 'outputs'. Il est toutefois essentiel d'étendre le questionnement plus loin dans la chaîne des résultats attendus, au-delà de la bonne mise en œuvre des activités. Il s'agirait idéalement d'examiner également les résultats au niveau des destinataires ou à un niveau plus large. En outre, près de la moitié des termes de référence des rapports examinés ne demandaient pas d'examiner la durabilité et l'impact des interventions. Il s'agit pourtant là de deux critères d'évaluation essentiels pour mesurer la mesure dans laquelle l'aide a réellement 'contribué au développement' des destinataires et pays partenaires. Il est souvent difficile de mesurer la durabilité par faute de recul dans le temps, et l'impact par faute de moyens ou de méthode adéquate. Néanmoins, il est essentiel d'examiner la durabilité et l'impact *potentiels* de l'intervention et les facteurs les favorisant. Ceci peut se baser sur diverses questions et *proxies* en la matière qui sont à la portée de la plupart des évaluations, comme par exemple : la prise en compte par l'intervention d'indicateurs de durabilité et d'impact, l'existence d'une stratégie de sortie, le degré d'appropriation par les destinataires, leur degré de satisfaction, leur capacité à faire perdurer les résultats, etc. En outre, il est recommandé de lancer des évaluations multi-acteurs sur l'impact conjoint de leur action. Ceci a été suggéré explicitement par plusieurs acteurs pour l'évaluation des activités

d'éducation au développement (sur le public belge). Cela aurait par ailleurs beaucoup de sens également sur certaines thématiques et dans certains pays du Sud dans lesquels plusieurs acteurs sont actifs, qu'ils soient belges, locaux ou internationaux.

- **R7.5 – Qualité des évaluateurs :** Cette étude a montré des lacunes importantes dans la conduite d'évaluations et la rédaction de rapports, par rapport à des standards de qualité largement admis en la matière (définition et description d'une méthodologie rigoureuse, constats solides et bien argumentés, etc.). Elle a également montré que les évaluations conduites par un bureau d'études ont dans l'ensemble eu une proportion de rapports de bonne qualité clairement plus importante que celles réalisées par des experts individuels. Lors de la définition des exigences sur l'équipe d'évaluation dans les termes de référence et au moment de la sélection de l'équipe, il est ainsi recommandé de s'assurer que l'équipe dispose d'une réelle maîtrise des concepts et outils de l'évaluation. L'équipe devrait aussi pouvoir disposer de l'encadrement nécessaire, notamment en termes de contrôle de qualité interne. L'évaluation devrait aussi bénéficier de l'esprit critique et de l'indépendance d'esprit nécessaires, ainsi idéalement que de bonnes pratiques d'autres organisations et de visions novatrices. Ceci peut être apporté par des personnes intégrées à l'équipe, mais également dans un comité de pilotage, ou lors d'activités d'évaluation telles que des entretiens avec des personnes ressources externes ou une revue de littérature. Des garanties formelles d'absence de conflit d'intérêt devraient par ailleurs être fournies pour les évaluations visant la reddition de compte.
- **R7.6 – Qualité du processus :** Cette méta-évaluation a montré que le processus d'évaluation était un des facteurs expliquant les différences de qualité. Il est recommandé à ce propos d'assurer un encadrement à l'évaluation visant à en favoriser la qualité. Ceci permettrait de faciliter l'accès des évaluateurs à l'information, d'effectuer un contrôle de qualité externe, d'attirer l'attention sur des dimensions particulières, etc. Cet encadrement serait typiquement effectué par un comité (appelé de pilotage, de concertation, d'accompagnement ou autre). Il inclurait idéalement des personnes connaissant bien le contexte de l'action ou du programme évalué, et, à l'opposé, des personnes en étant suffisamment éloignées que pour favoriser des réflexions '*out of the box*'. La composition d'un tel comité serait déterminée en fonction du commanditaire, des parties prenantes et de l'objectif de l'évaluation. Le rôle devra en être défini clairement ; il faudra notamment déterminer s'il est opportun de lui donner un pouvoir décisionnel. Un tel comité, s'il est adéquatement défini, devrait également augmenter la crédibilité de l'évaluation. Un dispositif tel que celui développé dans la R6 pourrait par ailleurs jouer un rôle dans l'encadrement des évaluations, y compris au sein d'un tel comité, à l'exemple de ce que fait le F3E en France.⁹⁶

⁹⁶ « Lorsque le F3E accompagne une étude, il suit avec la structure membre son déroulement, étape par étape : élaboration des termes de référence, identification et sélection de l'expertise externe, cadrage, suivi, restitutions, validation et valorisation de l'étude. Le F3E facilite le recours à l'expertise externe et organise le dialogue entre les différents acteurs. Éventuellement, il co-finance l'étude. » (Fonds pour la promotion des Études préalables, des Études transversales et des Évaluations (F3E), non daté. *Des ressources pour l'évaluation, des repères pour l'action.*).

Recommandation 8 : Conduire l'évaluation au meilleur moment
Le moment approprié pour évaluer devra être précautionneusement défini, en fonction de l'objectif poursuivi pour chaque type d'évaluation.
<i>Adressée à la DGD et aux ANG</i>
<i>Basée sur l'Encadré 8</i>

En fonction des objectifs définis par la DGD pour l'évaluation et des modalités envisagées, le moment le plus opportun pour conduire une évaluation devra être défini. Des évaluations visant à *rendre compte* devraient par exemple être conduites ex post, dès la fin des activités du programme pour bénéficier encore de la présence des partenaires et pour disposer du rapport sans tarder, ou justement avec une ou plusieurs années de recul pour pouvoir mieux mesurer l'impact et la durabilité. Des évaluations visant à *aider à la décision* devraient être réalisées à un moment utile pour la prise de décision par les ANG ou par la DGD, tout en veillant à disposer de suffisamment d'informations sur les résultats obtenus. Pour la *capitalisation et le partage des connaissances*, le moment de l'évaluation est souvent moins critique et dépendra surtout du contexte et du type de leçons à tirer.

Recommandation 9 : Renforcer la capacité interne en évaluation
La capacité en évaluation de la DGD devra être renforcée structurellement, d'un point de vue institutionnel, organisationnel et humain, en fonction de la répartition des responsabilités opérationnelles en matière d'évaluation entre les différentes parties.
<i>Adressée à la DGD</i>
<i>Basée sur la conclusion C2</i>

Comme décrit dans les Recommandations R1, R2 et R3, la répartition des responsabilités opérationnelles de la politique d'évaluation devra être définie entre les différents acteurs (la DGD, le SES, les acteurs non gouvernementaux ou des dispositifs externes existants ou ad hoc). Des moyens à la hauteur de leurs responsabilités devront leur être alloués afin de s'assurer de pouvoir atteindre l'effet escompté.

Il est de toute manière important que la DGD puisse disposer de suffisamment d'expertise pointue en évaluation, pour l'accompagner dans le développement de sa politique d'évaluation, en ce compris l'élaboration d'un système de rapportage sur les résultats, mais également de façon plus structurelle pour assurer le succès de la mise en œuvre de la politique d'évaluation et de ses modalités et dispositifs. Cette expertise pourrait dans un premier temps provenir du SES, d'expertise externe en évaluation (p.ex. d'experts individuels, de bureaux d'étude, d'ANG spécialisés en évaluation, ou de pairs), ou d'un mécanisme associant certaines de ces parties. Pour un accès plus structurel à de l'expertise en évaluation, la DGD pourrait par exemple disposer de ressources internes spécialisées, d'un accord/partenariat avec le SES, d'un accord/partenariat avec un ou plusieurs experts externes (cf. supra : bureaux, individus, ANG, pairs...), et/ou d'un dispositif externe associant en premier lieu la DGD et les ANG (comme celui recommandé dans la R6).

En parallèle, la DGD devrait s'assurer d'augmenter les capacités de son personnel en évaluation également de façon transversale. Il est notamment recommandé que l'ensemble des gestionnaires de programmes d'acteurs non gouvernementaux puisse suivre des formations en gestion d'évaluation. Ceci se ferait idéalement de façon conjointe avec les acteurs non gouvernementaux afin de favoriser une compréhension commune (cf. R6).

Il est par ailleurs recommandé de s'assurer que les processus d'évaluation soient intégrés clairement dans l'organisation, par exemple en termes d'instructions précises

sur l'utilisation des rapports, de classement systématique, de base de données en facilitant l'accès, etc.

Annexes

ANNEXE 1 -	CAHIER DES CHARGES
ANNEXE 2 -	CADRE RÉGLEMENTAIRE
ANNEXE 3 -	ÉCHANTILLON
ANNEXE 4 -	QUESTIONS D'ÉVALUATION
ANNEXE 5 -	MATRICE D'ÉVALUATION (ET RÉSULTATS)
ANNEXE 6 -	FICHE RAPPORT (ET RÉSULTATS)
ANNEXE 7 -	FICHE ACTEUR (ET RÉSULTATS)
ANNEXE 8 -	ENQUÊTE ANG (ET RÉSULTATS)
ANNEXE 9 -	ENQUÊTE ÉVALUATEURS (ET RÉSULTATS)
ANNEXE 10 -	ENQUÊTE DGD (ET RÉSULTATS)
ANNEXE 11 -	GRILLE DE CONCLUSIONS ET RECOMMANDATIONS RÉCURRENTES
ANNEXE 12 -	PRATIQUES INNOVANTES
ANNEXE 13 -	Liste des personnes rencontrées
ANNEXE 14 -	BIBLIOGRAPHIE

Les annexes de cette méta-évaluation sont reprises dans un volume séparé, disponible sur CD-ROM, auprès du Service de l'Évaluation spéciale ou sur le site

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

Réponse managériale de la DGD

Réponse managériale de la DGD à la méta-évaluation des programmes des acteurs non gouvernementaux (ANG)

Conclusion 1 : Absence d'une véritable politique d'évaluation

Pour la période 2008-2011, il n'y avait pas à la DGD de véritable politique d'évaluation fixant des objectifs, définissant des modalités et prévoyant les ressources nécessaires. Celle-ci s'est largement limitée à transférer la responsabilité de l'évaluation aux acteurs.

Recommandation 1 : Définir l'approche générale de la DGD en matière d'évaluation

La DGD doit revoir sa politique d'évaluation de la coopération non gouvernementale pour une reddition de compte crédible et pour améliorer l'efficacité de son aide. Elle doit pour cela élaborer une approche générale claire, définissant les objectifs poursuivis, les modalités, les dispositifs et un cadre réglementaire adéquats, ainsi qu'une allocation appropriée des responsabilités et des ressources.

(Adressée à la DGD, en collaboration avec les ANG.)

Accord. Le service Qualité et résultats a été créé au moment de la réforme de la DGD en 2012 entre autres dans le but de mettre en place une politique en matière de résultats au sein de la DGD. Une note stratégique est actuellement en cours d'élaboration et fera l'objet de consultations internes et externes. Une stratégie générale sur les résultats doit voir le jour d'ici la fin 2013 ; elle abordera également les questions du suivi et de l'évaluation. Cette note servira de base à la rédaction, par chaque direction, d'un plan opérationnel destiné à répondre aux problèmes que l'évaluation aura révélés. Cette note stratégique fixera des objectifs clairs en termes d'évaluation. Comme l'indiquent les évaluateurs, ceci se fera en concertation avec les ANG.

Conclusion 2 : Potentiel de l'évaluation peu utilisé par la DGD

La décision prise par la DGD de confier presque entièrement l'évaluation aux ANG l'a largement privée d'un outil clé pour rendre compte de l'utilisation des fonds publics qui leur étaient confiés, de capitaliser sur les leçons identifiées et de disposer de l'information générée pour alimenter sa prise de décision.

Accord partiel. Il est vrai que le potentiel n'est pas suffisamment utilisé. L'évaluation l'a pointé, les rapports d'évaluation ne sont pas systématiquement mis au service de la reddition de comptes et du développement des connaissances. La réponse à la première recommandation implique que la DGD améliorera cette situation en mettant en place une stratégie-cadre qui permettra de faire un meilleur usage des évaluations et de mettre davantage à profit le potentiel qu'offrent les rapports. Le choix de la partie qui se chargera de l'évaluation relève toutefois de critères d'ordre plutôt pratique. La DGD ne dispose pas des capacités requises pour assurer à elle seule l'organisation des

évaluations de l'ensemble des ANG, et souhaite par ailleurs responsabiliser les ANG en leur confiant leur propre évaluation. La DGD, en s'entourant des garanties nécessaires, souhaite développer un système qui rassemblerait les informations pertinentes pour la reddition de comptes, le développement des connaissances et la prise de décisions stratégiques. Ces garanties relèveront d'une part des critères de validation du système de maîtrise de l'organisation, une des conditions d'agrément des ANG, et d'autre part des critères de certification des systèmes d'évaluation comme prescrit par la loi.

Recommandation 2 : Développer un dispositif structuré pour rendre compte

La DGD devrait créer un dispositif structuré pour permettre une reddition de compte plus crédible et plus utile par l'évaluation (y compris les recommandations A2.1 à A2.5).

(Adressée à la DGD, en collaboration avec les ANG)

Accord. La DGD est consciente de la nécessité de mieux structurer la pratique de l'évaluation des organisations partenaires (et donc pas uniquement des ANG) si l'on souhaite disposer d'un rapportage des résultats qui soit à la fois plus représentatif et de meilleure qualité, deux conditions de crédibilité du système. La nouvelle loi relative à la coopération au développement fixe les prémisses de cette approche ; celle-ci est plus amplement explicitée dans le nouvel arrêté royal applicable aux ANG qui est en cours d'élaboration. L'amélioration de la qualité chez les ANG y est traduite par une mise en place d'exigences entre autres en matière de capacités en gestion axée sur les résultats chez les ANG dans le cadre de la validation du système de maîtrise de l'organisation (voir également la deuxième conclusion).

En outre, le nouvel Arrêté royal mentionnera de manière plus explicite les moments auxquels les ANG devront organiser leurs évaluations. L'usage que fera la DGD de ces évaluations sera examiné en détail dans la note stratégique à laquelle travaille actuellement la DGD. L'objectif est de consolider les résultats à un niveau plus élevé, dans la ligne de l'engagement pris par la DGD dans sa Réponse managériale à l'évaluation du système de rapportage des résultats de la DGD.

Recommandation 3 : Permettre à la DGD d'utiliser l'évaluation comme outil d'aide à la décision

La DGD devrait avoir la possibilité d'utiliser l'outil qu'est l'évaluation pour ses propres besoins d'aide à la décision.

(Adressée à la DGD)

Accord partiel. La DGD convient de la nécessité – exprimée dans la recommandation générale – de pouvoir utiliser les évaluations comme outils d'aide à la prise de décision. L'explication donnée plus loin dans le rapport sur la possibilité de commanditer ses propres évaluations ne cadre cependant pas avec la décision prise de fusionner le service de l'évaluation interne de la DGD avec le service de l'Évaluation spéciale (SES). La DGD entend anticiper son besoin d'évaluations en faisant figurer plus explicitement ses besoins dans la nouvelle réglementation en cours d'élaboration.

Pour ce qui est de la recommandation relative à la possibilité de commanditer des évaluations, la DGD marque son accord sur la suggestion de solliciter le SES. Ce fut déjà le cas en 2012. Il semble néanmoins difficile pour le SES de satisfaire rapidement des besoins nouveaux d'évaluations de la DGD au vu du planning déjà à sa charge et du cycle budgétaire connexe. La DGD a donc décidé qu'à l'avenir elle communiquerait un aperçu pluriannuel de ses besoins d'évaluations qui pourront ainsi être pris en compte dans le planning du SES.

Recommandation 9 : Renforcer la capacité interne en évaluation

La capacité en évaluation de la DGD devra être renforcée structurellement, d'un point de vue institutionnel, organisationnel et humain, en fonction de la répartition des responsabilités opérationnelles en matière d'évaluation entre les différentes parties.

(Adressée à la DGD)

Accord partiel. La DGD accepte d'investir dans une formation ad hoc à l'intention de son personnel, et envisage d'y associer les ANG dans l'optique de parvenir à une conception commune de la gestion axée sur les résultats. La DGD examinera également la possibilité d'instaurer des comités de pilotage conjoints, comme cela se passe déjà dans la coopération universitaire où l'expérience a démontré leur utilité.

La suggestion des évaluateurs de faire appel à l'expertise de structures externes est intéressante, mais compte tenu des restrictions budgétaires actuelles, la marge budgétaire à court terme sera limitée.

Conclusion 3: Évaluation utilisée par les ANG pour la prise de décision

Les ANG ont conduit des évaluations principalement dans l'optique d'améliorer la suite de leur programme. Ils l'ont nettement moins utilisée dans une perspective de rendre compte ou de partager les leçons avec de tierces parties.

Ne s'applique pas à la DGD.

Recommandation 4 : Laisser de la flexibilité pour l'aide à la décision des ANG

Une liberté d'action importante devra être donnée aux acteurs non gouvernementaux pour les évaluations qui visent à les aider dans leur propre prise de décision.

(Adressée à la DGD, en collaboration avec les ANG)

Accord partiel. La DGD accueille favorablement l'idée d'utiliser l'évaluation pour tirer des leçons et apporter des améliorations. Concernant la fréquence des évaluations, une bonne pratique, selon la DGD, consiste à organiser des évaluations intermédiaires de manière plus systématique, à l'exemple de ce qui se fait en matière de coopération bilatérale directe. Les évaluations s'avèrent utiles non seulement afin de remédier à des situations problématiques mais également afin de documenter des bonnes pratiques de programmes qui fonctionnent bien et desquels des leçons peuvent être tirées et utilisées dans d'autres programmes ou par d'autres acteurs.

Conclusion 4: Faible capacité à rendre compte

Les évaluations n'avaient qu'une faible capacité à rendre compte. En effet, elles n'ont souvent pas été conçues ni utilisées dans cette optique. Des problèmes de qualité et des questions sur l'indépendance affectaient par ailleurs couramment leur crédibilité. Le cadre réglementaire ne prévoyait pas non plus que les ANG dussent les transmettre à la DGD.

Accord. Cette conclusion vient confirmer d'autres constats opérés dans l'évaluation du système de rapportage des résultats de la DGD.

Conclusion 5: Peu de partage des leçons des évaluations

Les leçons des évaluations ont rarement été partagées au-delà du cadre de chaque programme.

Accord.

Recommandation 5 : Élaborer un dispositif actif pour le partage des leçons

La politique d'évaluation de la DGD devrait inclure un dispositif visant à favoriser un large partage des leçons.

(Adressée à la DGD et aux ANG.)

Accord partiel. Le défi que représente la gestion des connaissances dépasse le cadre des relations entre la DGD et les ANG. À l'instar des évaluateurs, la DGD est d'avis que l'idéal serait de voir évoluer la gestion des connaissances dans un climat le plus ouvert possible au partage des leçons. Or, l'évaluation a épinglé une initiative passée qui consistait à rassembler les leçons dans une banque de données et qui a échoué en raison d'une méfiance des acteurs par rapport aux risques de voir leurs expériences négatives utilisées à mauvais escient. Il semble donc préférable d'élargir le cadre de pareilles initiatives et d'organiser des séances d'apprentissage qui s'adressent à l'ensemble du secteur.

En outre, la réforme entreprise par la DGD en 2012 comptait parmi ses objectifs la volonté de renforcer le travail transversal et de développer des centres de connaissances. Par exemple, la direction thématique est désormais systématiquement impliquée dans l'évaluation des programmes des ANG, permettant ainsi de faire un usage optimal des connaissances sectorielles et thématiques présentes au sein de la DGD. Des experts externes sont également associés au processus.

Il convient enfin de signaler que la capitalisation sur les leçons tirées ne doit pas se limiter au contexte belge. Des organisations comme *International Initiative for Impact Evaluation* (3IE) par exemple réalisent des évaluations d'impact pouvant être utiles aussi bien à la DGD et aux ANG qu'à d'autres acteurs belges.

Conclusion 7: Qualité faible de nombreux rapports

La qualité des rapports d'évaluation pose problème. Même si elle est bonne dans l'ensemble pour une grande majorité des programmes d'universités et d'institutions scientifiques, elle est faible pour la moitié des programmes d'ONG et d'autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité et s'expliquent par une multitude de facteurs. Ces constats sur la qualité suggèrent que la culture et la maîtrise de l'évaluation restent un défi pour les ANG ainsi que pour nombre d'évaluateurs.

Accepté. Étant donné que la DGD n'a pas systématiquement mesuré la qualité des rapports d'évaluation, nos sources internes ne nous permettent pas de corroborer cette conclusion. Celle-ci va cependant dans le sens de ce qu'avait relevé la DGD concernant certains problèmes spécifiques comme la recherche de l'expertise idoine.

Recommandation 7 : Favoriser les facteurs de qualité

L'évaluation doit tenir compte des facteurs principaux affectant la qualité, tels que la qualité de la demande (objectifs, budget, etc.), des systèmes de suivi et d'évaluation performants au niveau des projets, la qualité des évaluateurs et la qualité du processus.

(Adressée à la DGD et aux ANG.)

Accord. Bien que la qualité des évaluations soit en grande partie de la responsabilité des ANG, la DGD mettra en place une stratégie en matière de résultats qui précisera clairement ce qui est attendu des évaluations, les objectifs des différents types d'évaluation, etc. Une formation sera en outre organisée à l'intention du service D3 qui sera axée d'une part sur la qualité des évaluations et d'autre part sur la gestion de la qualité. La DGD est disposée à organiser cette formation en collaboration avec les ANG.

La DGD espère que la qualité des évaluations progressera de manière indirecte grâce à l'exigence d'amélioration de la capacité d'évaluation au sein des ANG. Les exigences

posées aux ANG en matière de gestion axée sur les résultats pour la validation de leur système de maîtrise de l'organisation dans le cadre de l'obtention de leur agrément concourront à cette amélioration.

Conclusion 8: Problèmes dans la conception et la mise en place des programmes

De manière générale, les programmes des ANG ont été confrontés à des problèmes dans la conception de la stratégie d'intervention et dans la mise en place d'un système performant de mesure des résultats et impacts.

Recommandation 6 : Élaborer un dispositif fort pour le renforcement de capacités

a DGD et les acteurs non gouvernementaux doivent s'assurer que l'ensemble des acteurs aient des capacités solides en évaluation, en ce compris en termes de système interne de suivi et d'évaluation performant. (Adressée à la DGD et aux ANG.)

Accord. La DGD examinera la politique et les capacités d'évaluation des ANG dans le cadre de la procédure d'agrément. Ce sera une manière pour la DGD de s'assurer de la qualité des prestations des ANG dans ces domaines.

En outre, les gestionnaires de dossiers recevront une formation sur la qualité des rapports d'évaluation afin de pouvoir évaluer les rapports d'évaluation qu'ils reçoivent des ANG. Par l'intégration de cet aspect dans le dialogue entre la DGD et les ANG, la qualité pourra être améliorée.

Conclusion 9: Peu de jugement sur la contribution au développement

Les rapports d'évaluation ne permettent pas d'avoir un jugement global sur la contribution au développement par la coopération non gouvernementale.

Accord partiel. La DGD convient que les rapports d'évaluation en disent peu sur l'impact des projets/programmes réalisés. Il y a plusieurs raisons à cela. Idéalement, une évaluation d'impact se fait ex-post. Or, le financement accordé aux ANG porte sur la durée du projet ou du programme, excluant de ce fait la possibilité d'allouer un budget pour la réalisation d'évaluations ex-post de ce projet ou programme. En outre, le coût d'évaluations de grande ampleur étant très élevé, l'organisation systématique d'évaluations d'impact pour chaque projet/programme n'est pas réaliste. Il est vrai, enfin, que les évaluations existantes devraient se pencher davantage sur les potentiels en termes d'impact et de durabilité. L'évaluation d'impact qu'entreprend le SES de quatre projets de la coopération bilatérale directe fait cependant apparaître la nécessité d'appliquer des méthodologies solides si l'on souhaite récolter des informations de qualité sur l'impact de projets ou programmes. Il semble peu opportun, au vu du coût engendré par de pareilles méthodologies, de miser nos efforts sur les évaluations d'impact auprès des ANG. D'autre part, la DGD attend les recommandations de l'évaluation d'impact du SES pour décider des éventuelles améliorations méthodologiques à apporter.

Conclusion 10: Évaluations utiles pour les ANG et gestionnaires DGD

Les évaluations ont été considérées comme utiles à la prise de décision par la plupart des ANG. Les gestionnaires de la DGD les ont également largement appréciées, en particulier pour la valeur ajoutée qu'elles offraient par rapport au rapportage annuel par les ANG.

Accord.

Réponse managériale des ANG

Bruxelles, le 25 septembre 2013

Réponse managériale des Fédérations des ONG francophones et néerlandophones (ACODEV et ngo-federatie), des organisations syndicales (FGTB/IFSI, CGSLB/MSI, CSC/IEOI), de la coopération universitaire (CIUF-CUD et VLIR-UOS), de l'APEFE et du VVOB, l'ITG, le MRAC, la province d'Hainaut et le VVSG.

La 'méta-évaluation des programmes des acteurs non gouvernementaux' commanditée par le Service de l'Évaluation Spéciale est un rapport qui servira comme point de départ pour travailler à une meilleure politique et à de meilleures pratiques d'évaluation, que cela soit au niveau de l'acteur individuel ou au niveau collectif.

Par cette réponse managériale, nous voulons d'une part exprimer notre reconnaissance pour le processus constructif entre août 2012 et septembre 2013 et, d'autre part, formuler une appréciation sur les conclusions et les recommandations du rapport final. Cette appréciation ne représente en aucun cas une obligation individuelle de chaque acteur à traduire en pratique chaque recommandation.

Les acteurs soussignés ont pris la peine de rechercher un compromis sur une réponse managériale commune. Nous souhaitons ici mettre l'accent sur le fait qu'aujourd'hui il n'existe pas de constellation formelle qui puisse, au nom de tous les acteurs concernés (qui faisaient l'objet de la méta-évaluation), formuler un point de vue commun.

Il n'a pas été possible de suivre le format de réponse managériale qui nous a été remis par le Service de l'Évaluation Spéciale, en tant qu'acteurs qui sont dirigés par des instances très diverses :

- D'une part parce qu'il n'y a pas de soutien pour formuler un plan d'action commun pour les recommandations (il s'agit d'une responsabilité propre à chaque acteur)
- D'autre part parce que le format différait quelque peu des formulations originales telles que présentes dans le rapport. Nous nous sommes donc basés sur le texte original d'ADE (version d'août 2013)

Nous formulons ici une appréciation des conclusions générales et des recommandations spécifiques qui nous concernent en tant qu'acteurs de la coopération indirecte. Pour les recommandations dirigées à la DGD, nous formulons un commentaire ou un éclairage additionnel. Ces derniers ne remettent pas fondamentalement les recommandations en question.

Souscrivent à la réponse aux conclusions et recommandations de la méta-évaluation:



Voor/pour ABVV/ISVI - FGTB/IFSI,
Vera Van Nuffelen



Voor/pour KMMA – MRAC
Guido Gryseels



Voor/pour ACLVB/BIS
- CGSLB/MSI,
Vera Dos Santos Costa



Voor ngo-federatie,
Johan Cottenie



Pour ACODEV,
Etienne Van Parys



Pour la Province de Hainaut,
Chantal Vandermeiren



Voor/pour ACV/IIAV - CSC/IEOI,
Stijn Sintubin



Pour l'UVCW,
Isabelle Compagnie



Pour APEFE,
Stéphan Plumet



Voor VLIR-UOS,
Kristien Verbruggen



Pour/voor l'AVCB-VSGB,
Jean-Michel Reniers



Voor VVOB,
Bart Dewaele



Pour CIUF-CUD,
Murielle Andersson



Voor VVSG,
Bert Janssens



Voor/pour ITG,
Bruno Gryseels

Conclusion 1	Absence d'une véritable politique d'évaluation Pour la période 2008-2011, il n'y avait pas à la DGD de véritable politique d'évaluation fixant des objectifs, définissant des modalités et prévoyant les ressources nécessaires. Celle-ci s'est largement limitée à transférer la responsabilité de l'évaluation aux acteurs.
Management Response	Conclusion acceptée.
Conclusion 2	Potentiel de l'évaluation peu utilisé par la DGD La décision prise par la DGD de confier presque complètement l'évaluation aux ANG l'a largement privée d'un outil clé pour rendre compte de l'utilisation des fonds publics qui leur étaient confiés, de capitaliser sur les leçons identifiées et de disposer de l'information générée pour alimenter sa prise de décision.
Management Response	Conclusion acceptée partiellement. Les ANG rendent déjà des comptes sur l'utilisation des moyens et l'atteinte des résultats via le rapportage annuel et les rapports de fin de programme/projet. En principe, la DGD dispose donc de beaucoup d'information à exploiter en interne. Les ANG sont disposés à rechercher, avec la DGD, comment la manière actuelle de reddition des comptes peut être mieux utilisée par la DGD. Pour l'évaluation, l'accent devrait être mis davantage sur les autres objectifs (aide à la décision et capitalisation des connaissances) que sur la reddition des comptes.
Conclusion 3	Evaluation utilisée par les ANG pour la prise de décision Les ANG ont conduit des évaluations principalement dans l'optique d'améliorer la suite de leur programme. Ils l'ont nettement moins utilisée dans une perspective de rendre compte ou de partager les leçons avec de tierces parties.
Management Response	Conclusion acceptée.
Conclusion 4	Faible capacité à rendre compte Les évaluations n'avaient qu'une faible capacité à rendre compte. En effet, elles n'ont souvent pas été conçues ni utilisées dans cette optique. Des problèmes de qualité et des questions sur l'indépendance affectaient par ailleurs couramment leur crédibilité. Le cadre réglementaire ne prévoyait pas non plus que les ANG dussent les transmettre à la DGD.

Management Response:	Conclusion partiellement acceptée. “Les évaluations n’avaient qu’une faible capacité à rendre compte. En effet, elles n’ont souvent pas été conçues ni utilisées dans cette optique” : conclusion acceptée. “Des problèmes de qualité et des questions sur l’indépendance affectaient par ailleurs couramment leur crédibilité”: conclusion non acceptée parce que 1) Les pratiques d’évaluations varient fortement d’acteur à acteur et on ne peut donc pas généraliser, même au sein d’un groupe d’acteurs 2) L’importance de la crédibilité pour l’évaluation est fonction du type d’objectif poursuivi par elle. Comme le montre la C3, la plupart des ANG utilisent l’évaluation pour l’aide à la décision, pour laquelle la crédibilité telle que définie par les évaluateurs est sans doute moins cruciale. 3) L’indépendance n’est pas systématiquement synonyme de meilleure qualité. Cela aussi dépend des objectifs de l’évaluation : si l’accent est mis sur l’apprentissage, une évaluation par les pairs peut livrer une information utile. Bien entendu, si l’accent est mis sur la reddition de comptes, l’indépendance est plus importante.
Conclusion 5	Peu de partage des leçons des évaluations Les leçons des évaluations ont rarement été partagées au-delà du cadre de chaque programme.
Management Response	Conclusion acceptée. Elle découle principalement de la conclusion 3.
Conclusion 6	Progrès modeste des capacités des ANG en évaluation Les capacités en évaluation d’une majorité d’ANG se sont améliorées au cours de la période 2008-2011. Peu d’entre eux sont toutefois arrivés jusqu’au développement d’une politique interne d’évaluation et de guidance opérationnelle en la matière.
Management Response	Conclusion acceptée. (Pour beaucoup d’ANG, leur politique interne relève souvent d’un compromis avec la DGD. Il est d’ailleurs souvent difficile de mettre en œuvre une politique ambitieuse en raison des faibles budgets mobilisés à ce niveau.)

Conclusion 7	Qualité faible de nombreux rapports La qualité des rapports d'évaluation pose problème. Même si elle est bonne dans l'ensemble pour une grande majorité des programmes d'universités et d'institutions scientifiques, elle est faible pour la moitié des programmes d'ONG et d'autres partenaires. Les faiblesses observées sont liées à presque tous les critères de qualité et s'expliquent par une multitude de facteurs. Ces constats sur la qualité suggèrent que la culture et la maîtrise de l'évaluation restent un défi pour les ANG ainsi que pour nombre d'évaluateurs.
Management Response	Conclusion partiellement acceptée. La responsabilité de la faible qualité des évaluations est partagée : ANG (bonnes données de suivi, encadrement adéquat du processus d'évaluation), DGD (cadre budgétaire adapté, modalités flexibles), évaluateurs (méthodologie robuste, analyse critique). Il est parfois difficile d'obtenir des offres de bonne qualité pour les budgets disponibles.
Conclusion 8	Problèmes dans la conception et la mise en place des programmes De manière générale, les programmes des ANG ont été confrontés à des problèmes dans la conception de la stratégie d'intervention et dans la mise en place d'un système performant de mesure des résultats et impacts.
Management Response	Conclusion partiellement acceptée. Si on peut admettre le caractère perfectible de la conception et du suivi des interventions, ce constat n'est ni spécifique aux ANG ni spécifique à la Belgique. Par ailleurs, le cadre réglementaire influence directement la conception de la stratégie d'intervention. L'obligation d'utiliser des instruments spécifiques comme le cadre logique n'est pas satisfaisante dans certains contextes particuliers. Les ANG plaident en faveur d'une flexibilité pour mettre en œuvre des évaluations sur base de leur propre politique d'évaluation et sur base de leurs propres objectifs. (Il doit être ainsi possible, par exemple, de choisir entre des évaluations formatives ou sommatives) Les ANG sont prêts à étudier, avec la DGD, comme des systèmes de monitoring performants peuvent être développés pour mesurer l'impact (p.ex en mettant des moyens en commun) parce que nous sommes conscients que c'est impossible à faire au niveau d'une intervention individuelle.
Conclusion 9	Peu de jugement sur la contribution au développement Les rapports d'évaluation ne permettent pas d'avoir un jugement sur la contribution au développement de la coopération non gouvernementale.

Management Response	Conclusion acceptée. Les ANG estiment que c'est impossible à faire au niveau d'une intervention individuelle et y voient une responsabilité pour la DGD de rassembler l'information nécessaire pour pouvoir effectuer ce jugement sur la contribution au développement.
Conclusion 10	Évaluations utiles pour les ANG et gestionnaires DGD Les évaluations ont été considérées comme utiles à la prise de décision par la plupart des ANG. Les gestionnaires de la DGD les ont également largement appréciées, en particulier pour la valeur ajoutée qu'elles offraient par rapport au rapportage annuel par les ANG.
Management Response	Conclusion acceptée.

Recommandation 1	La DGD doit revoir sa politique d'évaluation de la coopération non gouvernementale pour une reddition de compte crédible et pour améliorer l'efficacité de son aide. Elle doit pour cela élaborer une approche générale claire, définissant les objectifs poursuivis, les modalités, les dispositifs et un cadre réglementaire adéquats, ainsi qu'une allocation appropriée des responsabilités et des ressources.
Management Response:	Recommandation adressée à la DGD <u>Commentaire des ANG :</u> La définition de cette politique conditionne la mise en œuvre de la plupart des recommandations qui suivent. Sans une vision claire des orientations de la DGD à cet égard, il est difficile pour les ANG de formuler un plan d'action spécifique. Il est donc important que le rôle et les responsabilités des ANG dans tout le processus soit défini. Les ANG sont prêts à soutenir toute initiative destinée à consolider un dispositif d'évaluation global et harmonisé (« socle commun ») pour la Coopération belge, auquel pourraient recourir la DGD et les ANG pour remplir leurs obligations en matière de reddition de compte et pour appuyer leurs décisions stratégiques. La DGD doit jouer un rôle d'appui. Concrètement, les ANG accueillent favorablement l'idée d'un cadre d'évaluation clairement structuré, cohérent, voire (partiellement) contraignant, pour autant qu'il soit élaboré avec les ANG et qu'il soit en ligne avec les systèmes de M&E des ANG et de leurs partenaires. Il est important que ce cadre laisse suffisamment de liberté pour la définition d'une politique propre à chaque ANG (p.ex. au niveau des objectifs (recommandation 4) et du timing (recommandation 8). Les ANG sont prêts à s'investir dans un exercice concret mené par la DGD pour aboutir à un tel cadre, pour autant que l'approche soit réellement participative et fondée sur le respect des attentes, de l'autonomie et des spécificités tant de la DGD que des ACNG. La question des ressources à accorder par la DGD pour la bonne mise en œuvre de cette politique est essentielle, en particulier, celles à prévoir pour permettre aux ANG d'assumer les responsabilités qui découlent de cette politique. En théorie beaucoup de choses sont possibles, mais en pratique, un budget doit être disponible pour chaque demande.
Recommandation 2	La DGD devrait créer un dispositif structuré pour permettre une reddition de compte plus crédible et plus utile par l'évaluation.
Management Response:	Recommandation adressée à la DGD <u>Commentaire des ANG :</u> Les ANG ne sont pas d'accord sur l'importance qui est mise sur la reddition des comptes comme objectif de l'évaluation, certainement pas si la mise en œuvre de ceci repose sur leurs épaules. Si, à leur niveau, l'évaluation a peu été utilisée comme moyen pour rendre des comptes, c'est parce que le système de reddition de comptes est basé sur d'autres sources comme le rapportage annuel. La reddition de compte sur les résultats est bien évidemment une responsabilité importante pour chaque acteur du développement. Les ANG sont disposés à alimenter la réflexion de la DGD sur la façon dont

	elle peut mieux tirer parti des outils actuels de reddition des comptes des ANG. Si le système actuel de reddition de comptes des ANG à la DGD basé sur un rapportage annuel / fin de projet/programme sur les résultats ne permettait pas à cette dernière de rendre des comptes, à son tour, de façon satisfaisante, il faudra analyser globalement les besoins en information de la DGD, les outils à mettre en œuvre et le partage des responsabilités en la matière. L'évaluation peut pallier certaines limites du système actuel. Toutefois, il n'est pas évident que ces évaluations devraient être mises en œuvre par les ANG eux-mêmes (surtout si l'on veut tirer parti des opportunités d'évaluations thématiques / pays, assurer une bonne crédibilité, approcher l'impact ou permettre la consolidation) ; et si c'était le cas, il est crucial d'identifier les gains d'efficacité qui pourront être réalisés par rapport à la situation actuelle, de telle sorte que l'évaluation ne devienne pas un nouveau mécanisme de contrôle qui puisse être utilisé contre les ANG.
Recommandation 3	La DGD devrait avoir la possibilité d'utiliser l'outil qu'est l'évaluation pour ses propres besoins d'aide à la décision.
Management Response:	Recommandation adressée à la DGD <u>Commentaire des ANG :</u> Les ANG sont sensibles au fait que la DGD ne se focalise pas uniquement sur les aspects négatifs relevés dans les évaluations pour sa prise de décision, mais que ces dernières soient un véritable support de dialogue stratégique entre ANG et DGD. Il serait d'ailleurs utile que la DGD rédige systématiquement une réponse managériale pour les remarques des évaluations qui s'adressent à elle. Il n'y a pas que les ANG qui devraient être tenus par les recommandations, la DGD également.
Recommandation 4	Une liberté d'action importante devra être donnée aux acteurs non gouvernementaux pour les évaluations qui visent à les aider dans leur propre prise de décision.
Management Response:	Recommandation adressée à la DGD <u>Commentaire des ANG :</u> Les ANG doivent disposer d'une marge de manœuvre suffisante pour définir et mettre en œuvre leur propre politique d'évaluation qui intègre leurs spécificités. La politique d'évaluation de la DGD doit tenir compte des pratiques existantes des ANG et de leurs partenaires et leur offrir un cadre souple soutenant la qualité des évaluations. En particulier, pour les évaluations axées sur la prise de décision, il est important que le cadre proposé par la DGD garantisse au maximum l'autonomie décisionnelle des ANG.
Recommandation 5	La politique d'évaluation de la DGD devrait inclure un dispositif visant à favoriser un large partage des leçons.
Management Response:	Les ANG sont d'avis qu'ils peuvent réaliser des progrès en matière de circulation des apprentissages tirés des évaluations. Une exigence de publication des évaluations pourrait sans doute y

	contribuer, mais elle doit être à tout le moins modulée en fonction des objectifs d'évaluation. Celle qui sert à l'aide à la décision ou à la capitalisation ne peuvent pas être rendues publiques de façon automatique. Par ailleurs, la publication des évaluations n'est certainement pas une condition suffisante pour identifier et faire circuler les apprentissages. Les ANG craignent que la publication des évaluations entraîne une image négative auprès du grand public. Il est toutefois bien entendu intéressant d'apprendre l'un de l'autre au sein du secteur (sans pour autant lier cela à la publication des évaluations). Les ANG sont néanmoins prêts à investir davantage dans le partage des leçons. Quelques tentatives ont déjà eu lieu (base de données des évaluations ONG) et il faudra tenir compte des leçons tirées. Plusieurs modalités pratiques plus ou moins ambitieuses sont envisageables et elles dépendront du cadre global de l'évaluation et des moyens que l'ensemble des acteurs (DGD et ANG) voudront bien y consacrer. Cela peut aller de - Prévoir que les termes de références requièrent systématiquement une section « leçons à partager » dans les rapports d'évaluation à - Organiser un mécanisme de type F3E pour le secteur qui exige des moyens spécifiques additionnels.
Recommandation 6	La DGD et les acteurs non gouvernementaux doivent s'assurer que l'ensemble des acteurs aient des capacités solides en évaluation, en ce compris en termes de système interne de suivi et d'évaluation performant.
Management Response:	Les ANG soutiennent cette recommandation. Si l'établissement d'exigences strictes relève de la responsabilité de la DGD, les ANG souhaitent être impliqués dans ce travail de sorte à ce que ces exigences soient partagées et soient perçues comme une opportunité d'amélioration de la qualité plutôt que comme une contrainte externe supplémentaire. En ce qui concerne l'appui aux capacités des ANG en matière d'évaluation, plusieurs approches sont possibles. Des efforts en formation ont déjà eu lieu par le passé. Il n'est pas certain qu'une approche unique soit opportune en raison de la spécificité propre à chaque catégorie d'ANG. Par ailleurs, les approches possibles dépendent aussi des moyens disponibles et de la priorité qui sera mise sur l'évaluation par la DGD. Les ANG feront, individuellement ou collectivement, des propositions concrètes à la DGD lorsque le cadre global sera connu.
Recommandation 7	L'évaluation doit tenir compte des facteurs principaux affectant la qualité, tels que la qualité de la demande (objectifs, budget, etc.), des systèmes de suivi et d'évaluation performants au niveau des projets, la qualité des évaluateurs et la qualité du processus.
Management Response:	Les ANG souscrivent globalement à la recommandation et favoriseront l'intégration, au sein de leurs politiques d'évaluation, des facteurs de qualité identifiés. Comme pour les autres recommandations, ils sont disposés à engager un dialogue stratégique avec la DGD pour l'identification de solutions pertinentes aux problèmes identifiés. La qualité dépend de deux facteurs : la politique d'évaluation (clarté des objectifs) et les budgets prévus. La

	responsabilité de la qualité ne peut donc pas uniquement être placée sur les épaules des ANG. Les ANG peuvent veiller à la qualité de plusieurs manières. En particulier : Pour R7.2, les ANG sont favorables à la généralisation à tous les ANG des lignes budgétaires spécifiques pour l'évaluation. Les moyens affectables à cette ligne budgétaire doivent être suffisants pour mener des évaluations de qualité. Pour R7.5, les ANG ne souscrivent pas à l'idée que les évaluations menées par des bureaux d'études seraient systématiquement de meilleure qualité que celles menées par des experts individuels. Par contre, l'offre d'évaluateurs de qualité en Belgique reste clairement un facteur limitant. Ceci est en partie dû aux budgets peu réalistes consacrés à l'évaluation par les acteurs. Dans ce cadre, il est crucial – comme le propose la recommandation – de laisser la porte ouverte à des pratiques innovantes ou d'autres formes d'évaluation (évaluation par les pairs, évaluation interne, évaluation par des consultants pensionnés ou des étudiants), lorsque l'objectif de l'évaluation le permet. Les ANG feront une proposition concrète à la DGD visant à diminuer l'impact de cette contrainte
Recommandation 8	Le moment approprié pour évaluer devra être précautionneusement défini, en fonction de l'objectif poursuivi pour chaque type d'évaluation.
Management Response:	Les ANG souscrivent à cette recommandation qui implique que dans certains cas, l'évaluation devra pouvoir être menée en dehors du cadre du financement du programme évalué. Des modalités pratiques devront être trouvées à cet effet.
Recommandation 9/ Aanbeveling 9:	La capacité en évaluation de la DGD devra être renforcée structurellement, d'un point de vue institutionnel, organisationnel et humain, en fonction de la répartition des responsabilités opérationnelles en matière d'évaluation entre les différentes parties.
Management Response:	Recommandation adressée à la DGD <u>Commentaire des ANG :</u> Une approche commune avec les ANG à ce niveau pourrait renforcer une culture partagée de l'évaluation entre ANG et DGD.



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