

Direction Générale des Relations Extérieures

Direction Générale du Développement

EuropeAid Office de Coopération

Unité Commune Évaluation

**MÉTHODOLOGIE D'ÉVALUATION DE
L'AIDE EXTÉRIEURE DE LA
COMMISSION EUROPÉENNE**



**BASES MÉTHODOLOGIQUES
D'ÉVALUATION**

VOLUME 1

Ni la Commission européenne ni aucune personne agissant au nom de la Commission n'est responsable de l'usage qui pourrait être fait des informations données ci-après.

De nombreuses autres informations sur l'évaluation de l'aide extérieure de la Commission européenne sont disponibles sur Internet Europa :

http://ec.europa.eu/europeaid/index_fr.htm

Luxembourg : Office des publications officielles des Communautés européennes

ISBN : 92-79-00682-7

© Communauté européenne, 2006

Reproduction autorisée moyennant mention de la source.

Imprimé en France, 2006.

Préface



L'amélioration constante des pratiques d'évaluation est essentielle pour la Commission européenne, particulièrement dans le domaine de l'assistance extérieure et du développement. Au souci de rendre compte aux contribuables européens de la bonne utilisation des fonds publics s'ajoute également la nécessité de tirer les leçons des interventions passées. La Commission européenne a pour cela besoin de méthodes d'évaluation crédibles afin de garantir la validité des conclusions et recommandations.

L'importance de l'évaluation des résultats des politiques régionales et sectorielles, des programmes et de l'efficacité de la programmation a été réaffirmée dans le cadre de la réforme de la gestion de l'aide extérieure lancée en 2000. L'évaluation est désormais un instrument essentiel pour le succès des politiques d'aide extérieure dans leur ensemble et elle est intégrée dans le cycle de programmation.

Entre 2002 et 2005, l'Unité commune Evaluation des directions générales EuropeAid, Développement et Relations Extérieures de la Commission européenne a développé une approche afin de répondre aux nouveaux besoins d'évaluation définis en 2000. Les résultats de ce travail méthodologique sont disponibles sur le site Web de la Commission et sont résumés dans quatre fascicules. Ces documents représentent un support utile pour mieux comprendre et maîtriser l'évaluation telle qu'elle a été mise en œuvre au sein de l'office de coopération EuropeAid. Je suis heureuse de constater que l'évaluation ainsi mise en œuvre constitue déjà une référence parmi les donateurs.

La démarche initiée par la Commission européenne est mise à la disposition du public et intéressera particulièrement l'ensemble des acteurs engagés dans le domaine de l'aide publique au développement et de la coopération avec les pays tiers.

Benita Ferrero Waldner

Synthèse des commentaires du panel d'experts sur le manuel d'évaluation d'EuropeAid

Impression générale concernant le manuel

Le projet de manuel correspond à un ensemble très complet de documents, décrivant de manière très détaillée le processus d'évaluation. Il reflète un effort engagé en vue de présenter de manière claire et structurée le processus d'évaluation et de fournir des indications répondant aux besoins des membres de l'organisation. Le manuel pourrait également être très utile à d'autres acteurs impliqués d'une manière ou d'une autre dans les processus d'évaluation d'EuropeAid, comme des consultants ou des représentants des autorités nationales de pays partenaires.

Le manuel est équilibré. Les éléments principaux se voient attribuer un poids égal et l'attention apportée à chaque aspect du processus d'évaluation est proportionnelle à son importance dans le processus. La plupart des sections constituant le manuel sont d'une longueur égale, d'une à deux pages. Le manuel traite souvent de concepts complexes avec une brièveté admirable. À la fin des structures hiérarchiques, les textes sont parfois plus longs, pouvant atteindre une quarantaine de pages (C'est le cas pour certaines sections relatives aux outils). Cela peut cependant se comprendre dans la mesure où les textes les plus longs se trouvent à la fin d'une ligne de recherche.

Le manuel s'adresse à différents types de lecteurs et peut être abordé de manière différente en fonction du lecteur et de ses centres d'intérêt. Cela s'avère utile et contribue à rendre le manuel flexible et évolutif.

Le contenu du manuel reflète pour une majeure partie un consensus émergeant au sein de la communauté de l'évaluation sur la manière de conduire une évaluation. À jour en 2005, il devrait demeurer utile à EuropeAid durant de nombreuses années.

Nos commentaires doivent être interprétés à la lueur d'un travail impressionnant qui reflète une parfaite compréhension de l'évaluation ainsi que des compétences substantielles pour retranscrire cette connaissance et la communiquer à d'autres au

travers du manuel et de la conception d'une source d'apprentissage organisationnel en ligne.

Questions de politique

Certains de nos commentaires se rapportent à la politique globale d'évaluation telle qu'elle est reflétée dans ce qui est écrit – ou ce qui n'est pas écrit – dans le manuel. Le choix d'amender le manuel pour tenir compte de nos observations dépendra de la position prise par EuropeAid concernant ces questions de fond.

Orientations et conseils pratiques concernant les évaluations conjointes

Il est souvent recommandé que les bailleurs de fonds joignent leurs forces en matière d'évaluation. Il existe de nombreuses raisons à cela, au premier rang desquelles figure le souhait d'alléger la charge du processus d'évaluation pour les pays en développement, dont le personnel responsable des projets et d'autres fonctionnaires consacrerait sinon trop de temps aux évaluations externes. De plus, les projets et programmes sont souvent financés et/ou mis en œuvre de manière conjointe, ce qui justifierait qu'ils soient évalués de manière conjointe. Il a été fortement recommandé par le Groupe de travail sur l'évaluation du CAD/OCDE que la coordination entre États membres s'améliore.

La coopération au développement est en train de changer et bien que l'appui budgétaire, l'assistance sectorielle et l'approche programmes se soient ajoutés à l'assistance par projets durant les deux ou trois dernières décennies, les approches dites « nouvelles » tendent à prendre de plus en plus d'importance. Ce sont également les approches de coopération les plus larges qui sont les plus adaptées à la coopération entre bailleurs de fonds et qui sont susceptibles de produire les plus grands effets dans les pays partenaires.

La coopération entre bailleurs suppose cependant des compromis. Dans la mesure où chaque bailleur a sa propre manière de codifier et de réguler le processus d'évaluation, une évaluation conjointe nécessiterait une importante adaptation mutuelle. Le manuel ne comporte aucun conseil ni orientation sur la manière dont les gestionnaires d'évaluation doivent aborder les évaluations conjointes. Dans le futur, il est probable que les gestionnaires d'évaluation auront de plus en plus à gérer des évaluations en collaboration avec des collègues d'autres institutions et ils doivent donc s'y adapter. C'est une tâche difficile, qui impose de

comprendre jusqu'à quel point certains aspects du processus d'évaluation sont « négociables » et, si oui, comment une collaboration peut s'établir.

À notre avis, le manuel gagnerait à être complété par une section sur l'évaluation conjointe et nous suggérons, en outre, que cette dernière ne soit pas décrite comme un problème ou une complication de plus, mais comme une opportunité de renforcer le processus d'évaluation, de réaliser des économies d'échelle et de rendre plus efficace la coopération avec le pays bénéficiaire.

Approche de l'évaluation participative

Un autre développement significatif dans le champ de la recherche en matière d'évaluation correspond à l'évaluation participative. Le concept est apparu au détour de travaux de recherche sur l'utilité et l'utilisation des résultats d'évaluation. Dans les cas où la fonction première de l'évaluation est formative, il est apparu particulièrement utile d'impliquer divers participants/parties prenantes dans les processus de définition de questions, de développement de méthodes, de collecte d'information et d'élaboration de conclusions. Plus les parties prenantes auront le sentiment de s'approprier les résultats intellectuels du processus, plus les chances seront grandes qu'elles agissent en prenant en compte les recommandations de l'évaluation. Le terme de parties prenantes fait ici référence aux organisations impliquées dans la coopération au développement telles que les ONG, les bureaux d'étude, les pouvoirs publics aux niveaux central et décentralisé, etc.

Lorsque d'autres acteurs prennent part au processus, il peut être nécessaire d'adapter ce dernier. Le groupe de référence qui est décrit dans le manuel peut, par exemple, être amené à remplir d'autres rôles et fonctions, et des organes supplémentaires pourraient également voir le jour. L'équipe d'évaluation peut également se voir confier d'autres rôles et avoir besoin de compétences supplémentaires, peut-être davantage en termes de facilitateurs du processus qu'en termes d'opérateurs. Comme pour l'évaluation conjointe, nous pensons que le manuel devrait contenir quelques conseils sur les stratégies d'évaluation participative, encourageant les gestionnaires d'évaluation à expérimenter les approches participatives et leur indiquant comment ajuster et adapter le processus afin de favoriser la participation d'autres acteurs.

Normes éthiques

La communauté de l'évaluation a consacré beaucoup d'attention à ce qui constitue la qualité en matière d'évaluation. Les « Normes d'Evaluation de Programmes » de 1994 (*The Joint Committee on Standards for Educational Evaluation. The Program Evaluation Standards. Thousand Oaks, CA: Sage Publications, Inc. 1994*) constituent l'une des publications pionnières les plus influentes sur le sujet. Le sujet de la qualité y est abordé selon quatre dimensions : l'utilité, la faisabilité, la convenance (*propriety*) et la précision. Le manuel que nous avons consulté traite en détail des aspects du processus d'évaluation qui concourent à la réalisation d'évaluations faisables, précises et utiles mais il contient peu d'éléments relatifs à l'éthique en évaluation.

Il y a de bonnes raisons de penser que les sujets éthiques se révéleront – à long terme – particulièrement aigus dans l'évaluation de la coopération au développement. Les relations de pouvoir asymétriques, la prédominance des modalités de pensée et d'action de type « donateur-bénéficiaire », les systèmes d'incitation souvent pervers qui entourent l'aide ainsi que les différences transculturelles contribuent à rendre l'évaluation de l'aide difficile et sujette à des choix éthiques complexes. Le manuel pourrait être complété par une section sur la signification des normes éthiques et la raison pour laquelle de telles normes sont importantes dans l'aide au développement. Le manuel pourrait également décrire quelques pièges éthiques auxquels les gestionnaires et les équipes d'évaluation sont susceptibles d'être confrontés – et dans lesquels il leur arrive de tomber.

Intégration avec des systèmes connexes

Pour autant que nous ayons compris, EuropeAid en tant qu'organisation applique un système de Gestion Basée sur les Résultats (*Results Based Management*). La manière dont ce système dans son ensemble, et la prise de décision en son sein, se rattachent au système d'évaluation n'est pas claire. Il existe une ambiguïté quant au fait que ces documents sont ou non conçus pour travailler dans la perspective d'une gestion basée sur les résultats (GBR). Certaines sections laissent supposer l'acceptation de la GBR comme cadre conceptuel pour ce document, tandis que le contenu d'autres parties suggère l'inverse - en particulier dans les sections sur les indicateurs et les niveaux cibles. Ce point appelle une attention particulière.

Nous pensons qu'une discussion est également nécessaire pour établir de manière explicite le lien rattachant les documents présentés ici aux fonctions de suivi et d'audit. Il n'y a pas de discussion sur les liens qui seraient nécessaires pour créer un système de suivi et d'évaluation dans un pays, pour les programmes, voire pour les projets. En fait, les documents établissent des distinctions entre l'évaluation, le suivi et l'audit, mais rien n'est dit des similitudes et des liens qui pourraient naître d'une plus grande coordination entre ces approches. Dans le passé, les évaluations de l'UE ont souvent été contraintes par l'utilité limitée des données de suivi. Il serait dommage de ne pas saisir cette occasion pour mieux coordonner ces activités.

Remarques finales

Au cœur de la tâche assignée au groupe d'experts figure une appréciation de la « qualité » de ces documents. La « qualité », cependant, est un concept notoirement ouvert aux interprétations, large ou étroite. Il existerait ainsi :

- (1) une définition étroite de la *qualité scientifique*
- (2) une définition plus large de la qualité comme utilité finale du document pour ses destinataires.

Au regard de la première définition (1), le travail réalisé par EuropeAid est bon. Certaines critiques peuvent être faites mais, en général, tant l'exhaustivité des textes que leur précision technique sont appréciables.

Par rapport à la deuxième acception (2), nous estimons qu'il reste du chemin à parcourir. La « motivation » est presque aussi importante que la « crédibilité » ou la « dissémination », or nous n'avons jusqu'ici vu que peu d'éléments propres à motiver les gestionnaires d'évaluation. Le document leur fournit des outils pour réaliser un travail mais il fait peu pour les convaincre qu'il s'agit d'un travail utile, créatif, intéressant et valant généralement la peine d'être fait. D'après ce que nous avons compris, beaucoup de gestionnaires d'évaluation dans les délégations devront assumer un grand nombre d'autres responsabilités, faisant courir le risque que l'évaluation ne devienne pour eux qu'une corvée supplémentaire – qu'il conviendrait de minimiser. Bien que le site Internet ne puisse par lui-même changer cette situation, il devrait au moins témoigner d'une certaine compréhension du problème.

Experts

Kim Forss, Andante, Stockholm

Kim Forss travaille pour Andante – tools for thinking AB. Son travail est dédié à la recherche et au conseil dans le champ de la recherche en évaluation, la recherche sur les politiques dans le domaine des relations internationales et le développement organisationnel. Il est l'auteur de communications relatives à la qualité et à l'utilisation des résultats d'évaluation dans la perspective d'une meilleure gestion démocratique.

John Mayne, Ottawa

John Mayne est un conseiller indépendant en matière de performance du secteur public. Auparavant, il a travaillé auprès du Bureau du Vérificateur Général du Canada et du secrétariat du Conseil du Trésor du Canada. Il a travaillé de manière approfondie sur l'appréciation de la qualité des évaluations. Il a également publié plusieurs ouvrages dans le domaine de l'évaluation de programme, dont notamment un livre international sur l'évaluation et la mesure de la performance.

Christopher Pollitt, Université de Rotterdam

Christopher Pollitt a une longue expérience de recherche et d'enseignement en évaluation et en gestion publique, d'abord à Brunel University (Royaume-Uni) et actuellement à l'Université de Rotterdam (Pays-Bas). Il a fait partie de plusieurs réseaux de recherche dans ces domaines en relation avec l'OCDE. Il a présidé la Société européenne d'évaluation. Il a publié une analyse critique d'une série de guides d'évaluations.

Ray Rist, Banque mondiale – OED, Washington

Ray Rist travaille actuellement comme conseiller méthodologique pour l'OED (Banque mondiale). Il a travaillé auparavant au sein du *General Accounting Office* des États-Unis. Il a une connaissance approfondie des pratiques d'évaluation au niveau mondial.

Pierre Spitz, INRA, Paris

Pierre Spitz travaille pour l'Institut national de la recherche agronomique (INRA) à Paris. Son domaine d'expertise concerne le développement rural. Il a également travaillé pour des agences des Nations unies au cours de la dernière décennie.

TABLE DES MATIÈRES

I. INTRODUCTION.....	13
II. PROCESSUS D'ÉVALUATION.....	15
1 Objet de l'évaluation	16
1.1 Projet, programme, stratégie.....	16
1.2 Secteurs, thèmes et aspects transversaux	18
1.3 Champ de l'évaluation	19
2 Moment de l'évaluation	21
2.1 Cas d'une intervention ponctuelle	22
2.2 Cas d'une suite d'interventions	22
3 Utilisation de l'évaluation	24
3.1 Les utilisateurs de l'évaluation	24
3.2 Les modes d'utilisation.....	26
3.3 La diffusion de l'évaluation	28
4 Les acteurs de l'évaluation et leurs rôles.....	30
4.1 Le gestionnaire de l'évaluation.....	30
4.2 Le groupe de référence	32
4.3 L'équipe d'évaluation.....	34
4.4 Les parties prenantes	35
4.5 Assistance qualité	35
III. METHODE D'ÉVALUATION.....	36
5 Stratégie d'intervention.....	36
5.1 Raison d'être de l'intervention.....	37
5.2 Reconstruire la logique d'intervention	38
5.3 Les politiques connexes	42
6 Questions d'évaluation	43
6.1 Utilité des questions	43

6.2	Origine des questions	44
6.3	Choisir les questions.....	46
6.4	Questions et critères d'évaluation	50
6.5	Préparer une question d'évaluation	52
7	Référentiel de jugement	55
7.1	Critères de jugement.....	55
7.2	Niveau cible (<i>target</i>)	57
7.3	Indicateurs	59
7.4	De la question à l'indicateur	62
8	Conception méthodologique	65
8.1	Introduction	65
8.2	Tableau de conception par question	66
8.3	Optimisation de la conception d'ensemble.....	70
8.4	Développement des outils	72
9	Collecte d'information	74
9.1	Le plan de collecte d'information	74
9.2	Problèmes fréquents et solutions.....	75
9.3	Fiabilité des données	77
10	Analyse.....	80
10.1	Stratégies d'analyse	80
10.2	Le déroulement de l'analyse.....	84
10.3	Validité de l'analyse.....	87
11	Jugement	89
11.1	Conclusions et leçons.....	89
11.2	Recommandations.....	92
12	Assurance qualité	93
12.1	Les règles du jeu.....	93
12.2	L'approbation des rapports	96

Glossaire.....	98
Disponible sur le site Web Europa	101
Liste des encadrés	101

I. INTRODUCTION

Vue d'ensemble

La Commission européenne a développé et formalisé la méthodologie d'évaluation de son aide extérieure en donnant la priorité aux résultats et aux impacts et en vue de maintenir ses évaluations à un niveau de qualité correspondant aux bonnes pratiques internationalement reconnues en la matière.

Dans le passé, l'évaluation de l'aide extérieure européenne s'est concentrée sur les projets et certains programmes. Cette aide méthodologique a vocation à faciliter le passage à une pratique d'évaluation orientée vers les programmes et les stratégies. Elle est principalement destinée :

- Aux gestionnaires des évaluations au siège de la Commission européenne et dans les délégations
- Aux équipes d'évaluation externes.

Elle est également mise à disposition de l'ensemble des partenaires de l'aide extérieure européenne ainsi que de la communauté professionnelle de l'évaluation.

Cette méthodologie d'évaluation est disponible en trois langues (anglais, espagnol et français) et sous deux formes optimisées respectivement pour la lecture et pour la navigation sur Internet.

La version Internet comprend de nombreux exemples et approfondissements et est disponible sur le site Web de la Commission européenne :

<http://ec.europa.eu/europeaid/evaluation/index.htm>

La version imprimée est composée de quatre volumes. Ce premier volume présente les concepts fondamentaux et leur articulation. Le second volume (Guide d'évaluation géographique et thématique) concerne l'évaluation de l'ensemble des actions communautaires à l'échelle d'un pays ou d'une région, ou de l'ensemble des actions relatives à un secteur ou à un thème à l'échelle globale. Le troisième volume (Guide d'évaluation de projet et de programme) concerne les grands projets, les projets pilotes, les programmes multi-pays et tout autre projet ou programme pour lequel une évaluation est requise. Le quatrième volume (Outils de l'évaluation) présente les principales techniques applicables pour la

structuration de l'évaluation, la collecte et l'analyse de l'information et l'aide à la formulation des jugements.

Bases méthodologiques

Ce volume présente l'essentiel de ce qu'il est souhaitable de savoir pour gérer ou réaliser une évaluation, avec une terminologie stabilisée dans trois langues (anglais, espagnol, français), et avec un vocabulaire contribuant aux efforts d'harmonisation entrepris dans le cadre du Comité d'aide au développement de l'OCDE (Cf. p 98).

Il offre des solutions complètes, concrètes et testées à tous les problèmes, y compris les plus difficiles tels que la clarification de la logique d'intervention, la rédaction des questions d'évaluation, la conception méthodologique ou l'analyse d'impact.

Il est structuré en deux grandes parties :

- Le processus : objet de l'évaluation, moment de l'évaluation, utilisation de l'évaluation, les acteurs de l'évaluation et leurs rôles.
- La méthode : la stratégie d'intervention, les questions d'évaluation (utilité, faisabilité, élaboration), les références de jugement (critères et indicateurs), la conception méthodologique, la collecte d'information, l'analyse, le jugement (conclusions, leçons et recommandations) et l'assurance qualité.

II. PROCESSUS D'ÉVALUATION

Qu'est-ce que l'évaluation ?

OCDE-CAD

- 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'efficacité 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.

DG Budget

- Jugement des interventions au regard de leurs résultats, impacts et des besoins à satisfaire (La Communication sur l'Evaluation : SEC (2000) 1051)

Les principales finalités de l'évaluation sont les suivantes :

- Contribuer à la conception des interventions et apporter des éléments nécessaires à la définition de priorités politiques
- Veiller à une utilisation efficiente des ressources
- Améliorer la qualité de l'intervention
- Rendre compte des résultats de l'intervention (*accountability*).

Les chapitres qui suivent sont consacrés au processus d'évaluation vu sous les quatre angles suivants : l'objet de l'évaluation (ce qui est évalué), le moment de l'évaluation, l'utilisation de l'évaluation, les acteurs et leurs rôles.

1 Objet de l'évaluation

RÉSUMÉ

Projet, programme, stratégie

L'intervention évaluée peut être simple, par exemple un projet qui vise à produire un effet donné sur un public donné. Au contraire, l'intervention peut être complexe, par exemple une stratégie qui produit ses effets à travers de multiples interventions. Selon le degré de complexité de l'intervention, les questions traitées et la méthode d'évaluation peuvent varier.

Secteurs et aspects transversaux

Les interventions publiques sont usuellement classées par secteurs ou domaines. Par ailleurs, une intervention donnée ou un ensemble d'interventions peuvent être évalués selon un aspect transversal.

Champ de l'évaluation

Le champ est l'objet des investigations de l'évaluation. Son périmètre est défini selon plusieurs dimensions : géographique, temporelle, institutionnelle et réglementaire.

Un second périmètre plus large précise les actions connexes et les éléments du contexte à prendre en compte.

1.1 Projet, programme, stratégie

Le terme « intervention » désigne ce qui est évalué, c'est-à-dire un ou plusieurs projet(s), programme(s), stratégie(s), ou tout autre forme d'aide extérieure.

Evaluation d'un projet

Un projet est une opération non divisible, délimitée en termes de calendrier et de budget et placée, le plus souvent, sous la responsabilité d'un seul opérateur.

Les projets les plus simples à évaluer sont homogènes (Cf. Encadré 1), c'est-à-dire conçus avec :

- Une activité principale mise en œuvre dans un seul contexte en direction d'un seul public
- Un seul résultat direct attendu pour ce public
- Un seul impact attendu au niveau plus global.

L'évaluation d'un projet simple est facilitée en ce sens que l'on se concentre sur une seule cause (l'activité principale) et un seul effet (résultat attendu pour le public visé).

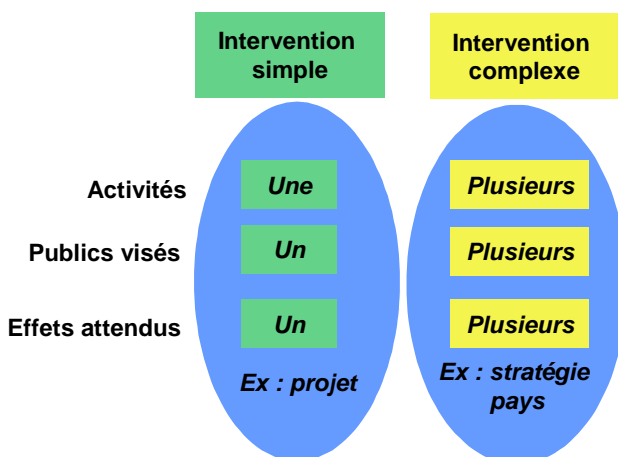
L'évaluation d'un projet privilégie souvent les questions qui intéressent les gestionnaires, les opérateurs et le public visé.

Evaluation d'un programme

Un programme est un ensemble d'interventions homogènes, regroupées pour atteindre des objectifs globaux. Il est délimité en termes de calendrier et de budget et placé, le plus souvent, sous la responsabilité d'un comité de pilotage.

Un programme homogène est constitué de projets parallèles et similaires. L'évaluation d'un programme homogène peut procéder par synthèse d'évaluations de projets ou par extrapolation à partir d'un échantillon de projets.

Encadré 1 – Interventions simples / complexes



Evaluation d'une intervention complexe

Les interventions complexes (par exemple une stratégie pays) regroupent des interventions hétérogènes mais complémentaires et mises au service d'un objectif commun. Elles sont conçues avec :

- Plusieurs activités (subventions, investissements directs, assistance technique, etc.) mises en œuvre dans plusieurs contextes (aires géographiques différentes, secteurs différents, ...) en direction de plusieurs publics
- De multiples résultats attendus pour ces différents publics
- Plusieurs impacts attendus au niveau plus global.

L'évaluation d'une intervention complexe n'est pas la somme ni la synthèse des évaluations de ses composantes. En effet, elle se concentre sur des questions qui sont pertinentes au niveau des décideurs politiques (effets de synergie, pertinence de la répartition des ressources entre les composantes, contribution à l'atteinte de l'objectif global).

1.2 Secteurs, thèmes et aspects transversaux

Les interventions publiques sont usuellement classées par secteurs. Par ailleurs, une intervention donnée ou un ensemble d'interventions peuvent être évalués selon un aspect transversal. Dans les deux cas, on parle aussi d'évaluation thématique.

Secteur

Les interventions sont classées dans un secteur ou un autre en fonction de la nature des activités et des réalisations.

Par exemple :

- Une aide à la formation des professeurs d'enseignement primaire est classée dans le secteur Education
- Un conseil offert à des groupements d'agriculteurs est classé dans le secteur Agriculture.

Aspects transversaux

Un aspect transversal (ou horizontal) se définit par rapport aux impacts et non par rapport aux réalisations, comme le montrent les deux exemples de l'encadré 2 :

Encadré 2 - Distinction entre secteurs et aspects transversaux

	Exemple 1	Exemple 2
Réalisation	Professeurs formés	Conseils donnés aux groupements d'agriculteurs
<i>Secteur</i>	<i>Education</i>	<i>Agriculture</i>
Impact	Les filles ont un meilleur accès à l'éducation	Nouvelles pratiques d'élevage limitant la désertification
<i>Aspect transversal</i>	<i>Genre</i>	<i>Environnement</i>

Thème

Le terme « thème » est quelquefois utilisé au sens de secteur et quelquefois au sens d'aspect transversal. Il vaut donc mieux le redéfinir à chaque utilisation.

Attention !

Un même terme peut désigner alternativement un secteur ou un aspect transversal. Par exemple, les droits de l'homme peuvent être :

- Un secteur, lorsqu'une évaluation porte sur un projet de formation du personnel des tribunaux.
- Un aspect transversal, lorsqu'une évaluation examine les effets d'un projet d'installation d'accès à Internet dans les zones rurales pauvres en ce qui concerne le nombre de signalements d'atteintes aux droits de l'homme.

1.3 Champ de l'évaluation

Le champ central de l'évaluation correspond au contour précis de l'intervention qui est soumise à l'évaluation. Le champ est défini

selon des dimensions variées telles que le territoire de référence, la période considérée, le cadre réglementaire, par exemple :

- Tous les financements alloués par l'Union européenne en Amérique latine depuis 1998
- Les actions de développement rural programmées au Rwanda au cours des quatre dernières années.

Un second périmètre précise les actions connexes et les éléments du contexte à prendre en compte, notamment pour l'analyse de la cohérence externe/complémentarité. Il s'agit du champ élargi, par opposition au champ central.

Délimitation du champ élargi

Les politiques connexes

- Les interventions des autorités nationales ou des autres bailleurs de fonds qui couvrent les mêmes territoires
- ... ou qui visent les mêmes publics
- ... ou qui traitent les mêmes problèmes

Les cofinancements

- Les interventions cofinancées avec les autorités nationales
- Les interventions cofinancées avec d'autres bailleurs de fonds.

Les autres éléments de contexte pertinents

Le champ central est précisé dans les termes de référence et le champ élargi dans le rapport de démarrage.

2 Moment de l'évaluation

RÉSUMÉ

Cas d'une intervention ponctuelle

Une évaluation peut être réalisée avant, pendant ou après l'intervention évaluée. Selon le moment, ses finalités et son utilisation seront différentes.

L'évaluation *ex ante* contribue à assurer la qualité de la conception.

L'évaluation en cours ou en fin de mise en œuvre vise à dégager les leçons des premières années de mise en œuvre de l'intervention et à ajuster le contenu de l'intervention en cours aux réalités du terrain et/ou à l'évolution du contexte.

L'évaluation *ex post* se préoccupe principalement de vérifier les impacts obtenus, et d'apprécier la durabilité des bénéfices de l'intervention.

Cas d'une suite d'interventions

Beaucoup d'interventions sont organisées par cycles successifs et présentent une relative continuité d'un cycle à l'autre.

Ces constats conduisent à choisir le moment de l'évaluation en considérant plusieurs cycles.

Il est important de se préoccuper du moment de l'évaluation :

- Pour optimiser les ressources allouées à l'évaluation en choisissant de la lancer au moment où elle aura la plus grande valeur ajoutée
- Pour répondre aux besoins des principaux utilisateurs de l'évaluation au meilleur moment
- Pour garantir qu'une masse critique de résultats et d'impacts s'est déjà matérialisée sur le terrain et se prête à une collecte d'information
- Pour éviter d'entrer en conflit avec des exercices concomitants de bilan ou d'audit.

2.1 Cas d'une intervention ponctuelle

2.1.1 Ex ante

L'évaluation *ex ante* est réalisée avant l'adoption ou avant la mise en œuvre de l'intervention et constitue une aide à sa conception. Elle vise à améliorer la conception de l'intervention dans la mesure où elle transpose les leçons des expériences passées dans le cadre de la nouvelle intervention.

2.1.2 En cours ou en fin de mise en œuvre

L'évaluation en cours ou en fin de mise en œuvre vise à dégager les leçons des premières années de mise en œuvre de l'intervention et à ajuster le contenu de l'intervention en cours aux réalités du terrain et/ou à l'évolution du contexte. Elle inclut souvent un bilan des réalisations achevées et une analyse des premiers résultats directs et impacts obtenus. Elle vise à améliorer l'intervention en cours et ses conclusions peuvent s'appuyer sur des observations de terrain.

2.1.3 Ex post

L'évaluation *ex post* est réalisée directement après ou longtemps après l'achèvement de la mise en œuvre. Elle se préoccupe principalement de vérifier les impacts obtenus, d'identifier et de juger les impacts inattendus et d'apprécier la durabilité des bénéfices de l'intervention.

Elle permet d'observer les changements réels sur le terrain et, si le délai n'est pas trop long, de les analyser pour estimer ceux qui sont attribuables à l'intervention.

L'évaluation *ex post* vise souvent à rendre compte aux institutions qui ont alloué les ressources. Elle aide également à transférer l'expérience acquise dans d'autres pays ou secteurs.

2.2 Cas d'une suite d'interventions

Beaucoup d'interventions sont organisées par cycles successifs et présentent une relative continuité d'une période à l'autre.

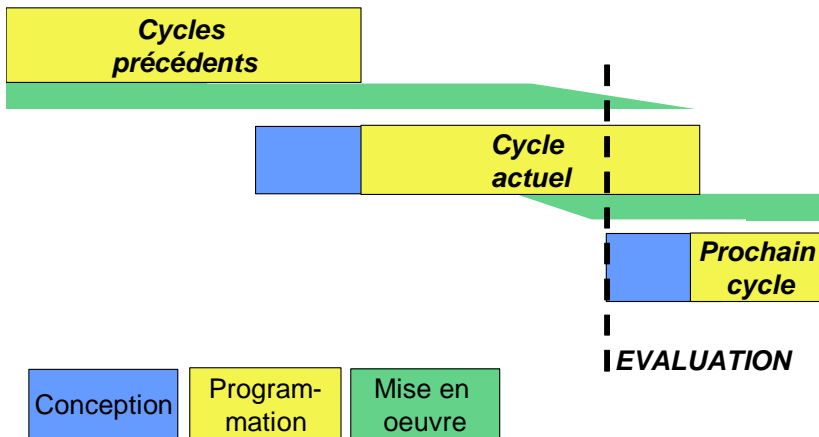
On constate souvent qu'un nouveau cycle débute alors même que l'intervention du cycle précédent n'a pas encore produit tous ses effets.

Ces constats conduisent à choisir le moment de l'évaluation en considérant plusieurs cycles :

- Pour tirer des leçons à partir d'une observation des effets des interventions des cycles précédents
- Pour nourrir la réflexion sur l'intervention du cycle prochain
- Pour observer la mise en œuvre de l'intervention du cycle en cours et permettre une rétroaction rapide.

Encadré 3 – Évaluation concernant plusieurs cycles

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009



3 Utilisation de l'évaluation

RÉSUMÉ

Les utilisateurs de l'évaluation

L'évaluation est destinée à des utilisateurs variés tels que les décideurs politiques et concepteurs des interventions, les gestionnaires et opérateurs chargés de la mise en œuvre, les partenaires, les institutions ayant attribué les financements et à qui des comptes doivent être rendus, les autorités publiques conduisant des interventions connexes ou similaires, les organisations de la société civile et les réseaux d'experts.

Les modes d'utilisation

Une évaluation peut être utilisée pour aider à préparer des décisions, pour formuler des jugements et/ou pour mieux connaître et comprendre l'intervention et ses effets.

La diffusion de l'évaluation

La diffusion vise à faire connaître le rapport final de l'évaluation, mais elle mobilise aussi toutes les autres façons de faire connaître les constats, les conclusions, les leçons apprises et les recommandations. La diffusion cible les services de la Commission, les institutions européennes, les partenaires externes, les réseaux d'experts, les médias et le grand public.

3.1 Les utilisateurs de l'évaluation

L'identification des utilisateurs est primordiale afin d'optimiser l'utilité de l'évaluation pour les différents partenaires, et en particulier :

- Pour s'assurer que l'évaluation répondra aux attentes des utilisateurs prévus, sous une forme et à un moment adaptés à leurs besoins
- Pour s'assurer que l'évaluation aura la crédibilité nécessaire vis-à-vis des utilisateurs prévus.

3.1.1 Décideurs politiques et concepteurs

Les décideurs politiques et les concepteurs utilisent l'évaluation pour préparer le lancement de nouvelles interventions, la réforme

d'interventions existantes, le choix d'orientations stratégiques, les décisions d'allocations de ressources budgétaires, humaines, etc.

Ils ont besoin d'une information :

- Directement utilisable dans le processus de décision
- Qui arrive à temps
- Qui répond à leurs questions de façon claire, succincte et fiable.

Ils s'intéressent aux enjeux stratégiques, à la cohérence externe et aux impacts globaux, qui constituent la finalité ultime de l'intervention.

3.1.2 Gestionnaires, partenaires et opérateurs

Les gestionnaires sont responsables de la mise en œuvre et du suivi de l'intervention, depuis les services centraux jusqu'au terrain, les plus proches du public étant les opérateurs. Les opérateurs de terrain appartiennent soit à la Commission européenne soit à des institutions partenaires qui partagent la responsabilité de la mise en œuvre.

Ils ont besoin d'une information qui arrive dès que possible pour corriger les écarts ou valider les changements.

Ils s'intéressent aux résultats directs de l'intervention, aux besoins et aux réactions du public visé, ainsi qu'aux interactions entre partenaires.

3.1.3 Autres acteurs

Les institutions qui ont attribué les financements attendent que des comptes leur soient rendus. Cela est vrai pour le Parlement et le Conseil des ministres, mais aussi pour tous les cofinanceurs. Les contribuables et les citoyens sont également destinataires de l'évaluation.

Les autorités publiques qui conduisent des interventions connexes ou similaires sont potentiellement utilisatrices de l'évaluation, notamment sous forme de transfert des leçons apprises. Il en va de même des réseaux d'experts concernés par l'intervention.

Finalement, l'évaluation est susceptible d'être utilisée par les acteurs de la société civile, en particulier ceux qui portent les intérêts des publics bénéficiaires.

3.2 Les modes d'utilisation

Une évaluation peut être utilisée :

- Pour aider à préparer des décisions
- Pour aider à formuler des jugements et informer l'opinion publique
- Pour mieux connaître et comprendre l'intervention et ses effets.

Ces trois modes d'utilisation ne sont pas exclusifs l'un de l'autre : il faut comprendre pour pouvoir juger et il faut juger pour pouvoir décider.

3.2.1 Aider à la décision

L'évaluation peut être conduite au bénéfice de ceux qui vont décider ou négocier un ajustement ou une réforme de l'intervention évaluée. Elle est alors utilisée pour corriger la mise en œuvre, pour concevoir le prochain cycle d'intervention ou pour redéfinir les orientations politiques.

L'évaluation peut aider à la décision de deux façons différentes :

- En formulant des conclusions de façon indépendante, puis en proposant des recommandations
- En favorisant l'implication des décideurs concernés, ou au moins leurs proches collaborateurs, en vue de créer un phénomène d'appropriation et de rétroaction directe en cours de processus.

L'aide à la décision prend différentes formes selon le contexte de la décision :

- Les recommandations peuvent s'adresser aux gestionnaires et/ou aux concepteurs de l'intervention
- Les recommandations peuvent s'adresser à l'ensemble des partenaires et cofinanceurs de l'intervention
- Enfin, l'évaluation peut être conçue comme une aide à la négociation et à la résolution de problèmes entre les parties prenantes de l'intervention, c'est-à-dire les institutions responsables, les groupes d'intérêts et les acteurs de la société civile.

3.2.2 Aider à la formulation des jugements

L'évaluation peut aider les utilisateurs à former leur opinion sur les mérites de l'intervention.

La formulation d'un jugement d'ensemble est particulièrement utile pour rendre des comptes. Dans ce cas, l'évaluation examine les mérites de l'intervention selon les différents points de vue. Elle répond aux questions qui sont importantes pour les institutions qui ont alloué les ressources. Le rapport est accessible au grand public. L'indépendance de l'évaluation et la transparence du jugement sont mises en valeur.

Dans ce cas, une attention particulière est apportée à la définition des critères de jugement (aussi appelés « critères d'appréciation raisonnée »). Le jugement lui-même n'est définitivement stabilisé qu'au moment de la remise du rapport final et de la discussion des conclusions. Utiliser une évaluation pour rendre des comptes suppose donc d'attendre la fin du processus.

3.2.3 Connaître et comprendre

Au-delà de l'aide à la décision et au jugement qui sont les deux formes d'utilisation généralement visées en priorité, l'évaluation permet aux utilisateurs d'améliorer leurs connaissances sur l'intervention et de mieux comprendre ses effets. Indirectement, elle contribue à transférer les connaissances acquises aux réseaux professionnels qui peuvent avoir un lien direct avec l'intervention évaluée.

A l'inverse de la rétroaction (*feedback*) qui concerne directement les responsables de l'intervention évaluée, le transfert des leçons est un processus indirect qui passe par les réseaux d'experts à l'intérieur et à l'extérieur de la Commission européenne.

La capitalisation des connaissances commence souvent durant le processus d'évaluation, par l'intermédiaire des experts sectoriels impliqués dans l'équipe d'évaluation et le groupe de référence. Cependant, le transfert des leçons apprises prend toute son importance après la remise du rapport final. La présentation de l'évaluation dans les réseaux spécialisés est une étape clé dans cette perspective. Elle prend la forme de séminaires ou d'articles techniques.

3.3 La diffusion de l'évaluation

La diffusion concerne le rapport final de l'évaluation, mais aussi toutes les autres façons de faire connaître les constats, les conclusions, les leçons apprises et les recommandations. La diffusion cible les services de la Commission, les institutions européennes, les partenaires externes, les réseaux d'experts, les médias et le grand public.

3.3.1 Les dispositions à prendre

La diffusion est prévue dès le stade des termes de référence, notamment en annonçant comment le rapport sera publié et quel sera le rôle de l'équipe d'évaluation dans cette phase.

Tout au long du processus, il importe de veiller à la qualité de l'évaluation et de ses produits (rapport, résumé, annexes) et d'apprécier formellement la qualité à la fin du processus.

La dernière réunion du groupe de référence est l'occasion d'identifier les principaux messages délivrés par l'évaluation et d'examiner comment atteindre les différents destinataires de ces messages.

Après approbation du rapport, le plan de communication est finalisé en choisissant les messages à mettre en valeur et les canaux d'information les plus adaptés pour chaque audience.

Enfin, il importe d'obtenir les coopérations nécessaires pour la mise en œuvre du plan de diffusion et de partager les rôles entre le gestionnaire de l'évaluation, l'équipe d'évaluation (vérifier que cela figure bien dans son mandat et que cela est spécifiquement mentionné dans les termes de référence) et les membres du groupe de référence et tout autre acteur pertinent.

3.3.2 Quels canaux pour la diffusion ?

Le rapport d'évaluation est en principe diffusé sur Internet de façon accessible à tous les publics. Une diffusion plus active est cependant entreprise pour des audiences spécifiques :

- Un résumé d'une à deux pages est préparé spécifiquement pour la hiérarchie du service qui a géré l'évaluation. Il pointe les principales conclusions et recommandations. Il est accompagné du rapport d'évaluation.

- Le rapport et ses annexes et/ou son résumé sont envoyés aux services concernés et aux partenaires.
- Un résumé est également placé sur d'autres sites Intranet pertinents, accompagné d'un lien vers le rapport.
- Un résumé est diffusé par Internet, à destination de la communauté internationale de l'aide au développement. Il pointe les leçons apprises dans la mesure où elles sont transférables.
- Un ou plusieurs articles peuvent être rédigés à destination du grand public ou de réseaux spécialisés.
- Enfin, le rapport peut être présenté au cours de réunions, ateliers ou séminaires.

Conseils pratiques pour une présentation

Une présentation courte (10 à 20 minutes) suffit pour exposer l'essentiel des messages résultant de l'évaluation mais un temps plus important doit être laissé aux questions (20 à 40 minutes).

La présentation aborde les points suivants :

- L'intervention évaluée
- La méthode d'évaluation
- Les messages résultant de l'évaluation
- Les limites méthodologiques de ces messages le cas échéant.

Sur le site Web : exemples de résumés et articles sur une évaluation

4 Les acteurs de l'évaluation et leurs rôles

RÉSUMÉ

Gestionnaire de l'évaluation

Le gestionnaire de l'évaluation (*evaluation manager*) est un membre du service en charge de l'évaluation. Il gère tout le processus au nom de la Commission européenne.

Groupe de référence

Le groupe de référence est présidé par le gestionnaire de l'évaluation et assiste celui-ci dans la commande et la supervision de l'évaluation. Il est composé de membres des services de la Commission européenne et de représentants du pays partenaire quand cela est possible.

Equipe d'évaluation

L'équipe d'évaluation est responsable de la collecte et de l'analyse des données ainsi que de la formulation des jugements en réponse aux questions d'évaluation. Elle rédige le rapport d'évaluation et en est responsable. Elle présente périodiquement ses travaux au groupe de référence et au gestionnaire de l'évaluation et réagit à leurs commentaires.

Parties prenantes

Les parties prenantes (*stakeholders*) sont les individus, les groupes ou les organisations qui ont des intérêts directs ou indirects dans l'intervention évaluée et dans l'évaluation elle-même. Ils peuvent être affectés ou non par l'intervention.

4.1 Le gestionnaire de l'évaluation

Le gestionnaire de l'évaluation (*evaluation manager*) est un membre du service qui est en charge de l'évaluation. Il gère tout le processus du début à la fin au nom du commanditaire de l'évaluation.

Il est désigné :

- Pour assurer la cohérence du processus d'évaluation, depuis les termes de référence jusqu'à la diffusion du rapport et au suivi des recommandations
- Pour être le référent administratif et coordonner l'activité des différents acteurs (groupe de référence et équipe d'évaluation)
- Pour organiser, superviser et vérifier la qualité du travail aux différentes phases
- Pour être garant du bon déroulement de l'évaluation.

Check-list : rôle du gestionnaire

Le gestionnaire de l'évaluation

- Constitue le groupe de référence après consultation des chefs des services concernés
- Rédige les termes de référence après consultation des membres du groupe de référence
- Procède à l'engagement de l'équipe externe d'évaluation après consultation du groupe de référence
- Aide l'équipe d'évaluation à accéder aux informations et aux informateurs
- Organise la discussion des conclusions et recommandations en interaction avec le groupe de référence et l'équipe d'évaluation
- Procède à l'appréciation de la qualité méthodologique des rapports, avec un appui pour seconde vérification
- Diffuse les résultats de l'évaluation auprès des différents acteurs concernés
- Suit la prise en compte des recommandations / leçons en interaction avec les services concernés.

Il est recommandé de prévoir que le gestionnaire de l'évaluation se fasse aider dans son travail quotidien par un suppléant, qui puisse le remplacer en cas d'absence.

4.2 Le groupe de référence

Le groupe de référence est une interface entre le gestionnaire et l'équipe d'évaluation. Il permet l'expression de la variété des points de vue portés sur l'intervention évaluée.

Check-list : rôle du groupe de référence

Les membres du groupe de référence :

- Discutent et commentent les termes de référence préparés par le gestionnaire de l'évaluation
- Rassemblent et résument les points de vue des services de la Commission et jouent un rôle d'interface entre l'équipe d'évaluation et les services, complétant ainsi les contacts bilatéraux
- S'assurent que l'équipe d'évaluation a accès à toutes les sources d'information et de documentation sur les activités mises en œuvre et les consulte
- Valident les questions d'évaluation
- Discutent et commentent les notes et rapports produits par l'équipe d'évaluation. Les commentaires de chaque membre du groupe de référence sont rassemblés par le gestionnaire d'évaluation, puis transmis à l'équipe d'évaluation
- Assistent le processus de rétroaction (*feedback*) à partir des constats, conclusions et recommandations de l'évaluation.

Si l'évaluation est menée à partir du siège, le groupe comprend :

- Les services de la Commission concernés par l'intervention, y compris la Délégation concernée dans le cas d'une évaluation au niveau pays/région
- Des spécialistes d'un secteur ou d'une question transversale au sein de la Commission
- L'ambassade du pays partenaire dans le cas d'une évaluation pays.

Si l'évaluation est menée à partir du pays partenaire, le groupe peut être élargi :

- Aux services du Gouvernement
- A une sélection de partenaires de développement

- A une sélection d'organisations de la société civile
- A des experts.

L'expérience a montré qu'il est préférable que le groupe soit limité en nombre (10 à 12 membres) pour que son fonctionnement soit efficace.

Après avoir identifié les services, institutions ou organisations à inviter au groupe de référence, le chef du service gestionnaire de l'évaluation leur envoie une note d'invitation pour :

- Annoncer officiellement le lancement de l'évaluation
- Demander la désignation d'une personne comme membre du groupe de référence.

Encadré 4 – Calendrier et rôle du groupe de référence

Etape	Réunion	Rôle
Termes de référence		Commentaires
Encadrement de l'équipe d'évaluation		Commentaires
Questions d'évaluation	1	Validation
Rapport de démarrage (<i>inception</i>)		Commentaires
Rapport de documentation (<i>desk</i>)	2	Discussion Commentaires
Débriefing de la phase de terrain	3	Discussion
Projet de rapport final, conclusions et recommandations	4	Discussion Commentaires

Sur le site Web : check-list pour organiser et animer une réunion du groupe de référence

4.3 L'équipe d'évaluation

L'équipe d'évaluation est responsable de la conception méthodologique, de la collecte et de l'analyse des données ainsi que du rapport, y compris la formulation des jugements (appelés également appréciations raisonnées) et de la rédaction des conclusions et recommandations. Elle interagit avec le groupe de référence et le gestionnaire de l'évaluation. Elle apporte ses services au commanditaire dans le cadre d'un contrat.

Les membres de l'équipe d'évaluation sont indépendants des organisations qui ont participé à la conception et à la mise en œuvre de l'intervention évaluée.

Critères pour engager l'équipe d'évaluation

- Connaissance des instruments d'aide au développement évalués
- Connaissance du champ de l'évaluation (secteur, pays, etc.)
- Présence d'experts et consultants locaux
- Compétence managériale du leader de l'équipe (organisation, coordination, dialogue avec le client, etc.)
- Indépendance de l'équipe vis-à-vis de l'intervention évaluée et absence de conflit d'intérêt
- Compétence d'évaluation et en particulier capacité à maîtriser la formulation des jugements, à rédiger des conclusions et recommandations et à réaliser des synthèses
- Maîtrise des outils de collecte et d'analyse de l'information
- Diversité des membres de l'équipe en terme de profils et d'expertise
- Prix.

Le gestionnaire fixe le poids relatif des critères et le précise dans les termes de référence.

Associer des consultants locaux

Associer autant que possible des consultants locaux permet d'aider au développement d'une compétence locale (construction de capacités nationales d'évaluation) et de bénéficier de leur bonne connaissance du terrain.

4.4 Les parties prenantes

Les parties prenantes (*stakeholders*) sont les individus, les groupes ou les organisations qui ont des responsabilités et/ou des intérêts directs ou indirects dans l'intervention évaluée. Ils peuvent être affectés ou non par l'intervention.

L'association des parties prenantes à l'évaluation est utile pour :

- Tenir compte de leurs points de vue
- Profiter de leur expertise et des informations qu'ils détiennent
- Les encourager à utiliser l'évaluation.

4.5 Assistance qualité

Il revient au gestionnaire de l'évaluation de garantir que les conclusions ont des bases méthodologiques et factuelles solides.

Au sein de l'Unité Evaluation au siège de la Commission européenne, les gestionnaires ont la formation et la capacité nécessaires pour apprécier la qualité méthodologique des rapports. Toutes les évaluations sont suivies par deux membres de l'Unité et tous deux participent à l'appréciation de qualité.

Dans le cas d'évaluations déconcentrées, le gestionnaire utilise les critères de qualité proposés sur le site Web ainsi que les indications données pour remplir la grille d'appréciation de qualité, si besoin avec l'appui de l'Unité Evaluation.

Sur le site Web : grille d'appréciation de la qualité et indications pour remplir la grille

III. METHODE D'ÉVALUATION

Les chapitres qui suivent sont consacrés à huit composantes clés de la méthode d'évaluation présentées de façon logique (et non pas chronologique).

Sont donc détaillés : la stratégie d'intervention, les questions d'évaluation (utilité, faisabilité, élaboration), les références de jugement (critères et indicateurs), la conception méthodologique, la collecte d'information, l'analyse, le jugement (conclusions, leçons et recommandations) et l'assurance qualité.

5 Stratégie d'intervention

RÉSUMÉ

La première étape du processus d'évaluation consiste à examiner la stratégie d'intervention, c'est-à-dire sa raison d'être, sa logique et sa place par rapport aux politiques connexes.

Raison d'être de l'intervention

La raison d'être d'une intervention est de traiter des besoins, des problèmes ou des enjeux qui sont considérés comme prioritaires dans le contexte concerné et qui ne peuvent pas être traités plus efficacement par une option alternative. Au stade de l'évaluation, il convient d'en rappeler les principaux éléments ou de les reconstruire si les documents de programmation sont imprécis.

Logique d'intervention

La logique d'intervention identifie les activités, les réalisations, les résultats et les différents niveaux d'impacts attendus. Il est nécessaire de rappeler ou de reconstruire la logique de l'intervention pour comprendre quels effets sont attendus et choisir les questions d'évaluation à poser.

Les politiques connexes

Il convient également d'identifier les principales interventions connexes (autres politiques de la Commission européenne, du pays partenaire, des autres bailleurs) et de situer l'intervention évaluée par rapport à elles. Il s'agit de comprendre où se trouvent les complémentarités, les synergies potentielles, les risques de duplication et les besoins de coordination.

5.1 Raison d'être de l'intervention

La raison d'être d'une intervention est de traiter des besoins, des problèmes ou des enjeux qui sont considérés comme prioritaires dans un contexte donné et qui ne peuvent pas être traités plus efficacement par une option alternative.

C'est au stade de la programmation qu'il faut justifier la raison d'être de l'intervention. Au stade de l'évaluation, il convient d'en rappeler les principaux éléments ou de les reconstruire si les documents de programmation sont imprécis.

L'évaluation identifie les raisons pour lesquelles les priorités ont été choisies, par exemple : priorités de la politique dans laquelle l'intervention prend place, urgence des besoins à satisfaire, avantages comparatifs par rapport à des options alternatives.

L'analyse procède en deux étapes comme suit :

Diagnostic des problèmes et réponses

Le rapport d'évaluation expose succinctement les éléments suivants :

- Contexte de l'intervention au moment où elle a été initiée
- Principaux problèmes (besoins, enjeux) diagnostiqués
- Pourquoi certaines options stratégiques ont-elles été choisies plutôt que d'autres ?

Certaines questions d'évaluation peuvent porter sur les problèmes traités par l'intervention, par exemple :

- Existait-il des options alternatives pour résoudre les problèmes identifiés ?
- Entre la date de la mise en place de l'intervention et celle de l'évaluation, les problèmes qui ont justifié l'intervention ont-ils changé de nature ou d'ampleur ?

Justification de l'intervention

Lorsque cela est pertinent, le rapport d'évaluation examine si les besoins, problèmes et enjeux pourraient :

- Se résoudre spontanément par des initiatives privées
- Se résoudre de façon plus efficace par d'autres initiatives publiques.

Recommandations

Les documents officiels privilégient souvent les options stratégiques qui ont été retenues *in fine*. Il est utile de chercher à comprendre quelles étaient les autres options.

5.2 Reconstruire la logique d'intervention

La logique d'intervention est constituée de l'ensemble des activités mises en œuvre, des effets attendus (réalisations, résultats et impacts comme indiqué dans l'encadré 8) ainsi que des hypothèses qui expliquent comment les activités vont conduire aux effets dans le contexte de l'intervention.

Reconstruire la logique d'intervention est nécessaire pour :

- Aider à clarifier les objectifs et à les traduire en une hiérarchie d'effets attendus de façon à permettre leur évaluation
- Suggérer des questions d'évaluation à propos de ces effets
- Aider à juger la cohérence interne de l'intervention.

La logique d'intervention peut être « fidèle » aux documents de programmation et aux documents fondateurs de la politique dans laquelle s'inscrit l'intervention. Dans ce cas, les effets attendus sont déduits des objectifs affichés dans les documents officiels.

Lorsque l'on reconstruit la logique d'intervention au moment de l'évaluation, on peut prendre en compte des effets qui sont implicitement attendus mais qui n'étaient pas évoqués dans les documents initiaux. On s'écarte alors de la logique « fidèle » et cela doit être précisé. L'approche « fidèle » n'est pertinente que quand les objectifs sont exprimés assez précisément et d'une manière vérifiable.

Il est fréquent que la logique d'une intervention évolue dans le temps. Le cas échéant, cela justifie de reconstruire plusieurs logiques correspondant à des périodes successives.

La logique d'intervention est une simplification utile de la réalité, mais il faut toujours rester conscient de la complexité du monde réel. En plus de la reconstruction de la logique d'intervention, il est donc utile d'identifier les principaux facteurs externes qui conditionnent ou qui contraignent la mise en œuvre et les effets. Il faut aussi garder en mémoire le fait que les explications causales réelles se révèlent souvent plus complexes que les hypothèses de départ.

5.2.1 Comment la reconstituer ?

La reconstitution de la logique d'intervention procède comme suit :

- Collecter et analyser les documents officiels fondateurs de l'intervention
- Identifier les principales activités
- Identifier les objectifs
- Traduire les objectifs sous forme de résultats et impacts attendus
- Connecter les activités aux impacts attendus en reconstituant des relations de cause à effet
- Vérifier que les relations causales sont logiques, c'est-à-dire qu'elles peuvent être considérées comme plausibles à la lumière des connaissances disponibles
- Discuter la logique reconstruite avec quelques informateurs clés (concepteurs et gestionnaires) et avec des experts du domaine / pays concerné
- Présenter et discuter la logique d'intervention en réunion de groupe de référence.

5.2.2 Les présentations les plus courantes

Cadre logique

La technique du cadre logique consiste à produire une présentation sous la forme d'une matrice qui spécifie les objectifs (première colonne), la façon dont ils devront être vérifiés (deuxième colonne : indicateurs et troisième colonne : sources), et les

principales hypothèses sur les facteurs externes, les conditions et les risques (quatrième colonne).

Les lignes du cadre logique reprennent une présentation simplifiée de la logique d'intervention, en seulement quatre niveaux : activités, résultats, finalité (*purpose* = avantage direct pour le public visé) et objectif global.

Cette représentation convient pour une intervention simple, comme un projet ou un programme bien ciblé. En revanche, elle ne peut saisir correctement la complexité d'une intervention hétérogène comme un programme de développement rural intégré, une stratégie pays ou l'appui à une politique sectorielle.

En raison de sa simplicité, le cadre logique permet de détailler les indicateurs qui devront être utilisés à chaque ligne, ainsi que les facteurs externes.

Diagramme des objectifs

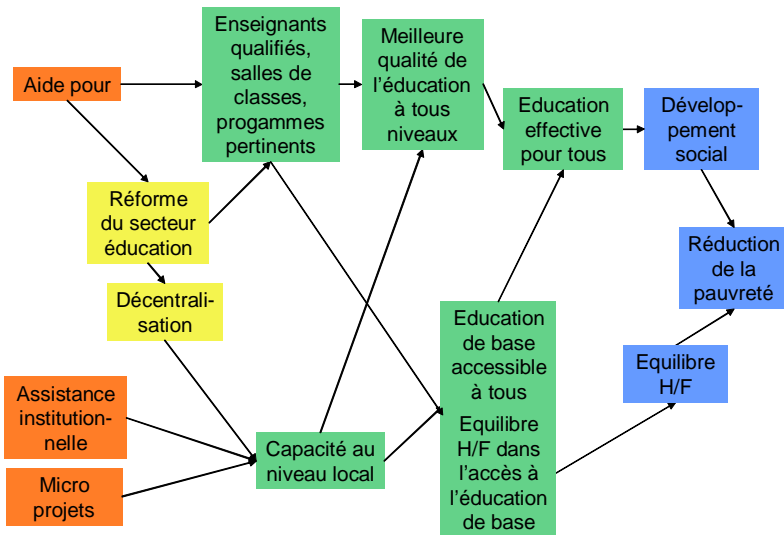
Le diagramme des objectifs (Cf. Volume 4) procède par identification des objectifs officiellement affichés et une présentation graphique montrant les enchaînements logiques entre objectifs, depuis les plus opérationnels jusqu'aux plus globaux. La logique d'intervention est représentée sous forme de boîtes et de flèches.

Une forme particulière de représentation est l'arbre d'objectifs. Elle s'applique dans le cas où chaque objectif contribue à un seul objectif plus global.

Diagramme des effets attendus

Le diagramme des effets attendus (Cf. Volume 4) est proche de la technique précédente. Il procède également par identification des objectifs officiellement affichés, mais les objectifs sont traduits sous forme d'effets attendus avant d'être présentés sous forme de diagramme. La traduction des objectifs en effets attendus permet de travailler avec des concepts plus concrets et plus facilement vérifiables.

Encadré 5 – Diagramme des effets attendus



Contrairement au cadre logique, le diagramme peut avoir des chaînes d'objectifs aussi longues que nécessaire. Il permet aussi de mettre en évidence des synergies et des relations complexes entre objectifs. Cette représentation convient bien pour des interventions complexes comme les programmes intégrés, les stratégies pays ou les politiques sectorielles.

Cependant, le diagramme ne permet pas de montrer les facteurs externes, les conditions et les risques comme dans la dernière colonne d'un cadre logique. Ces éléments sont précisés au stade de la conception méthodologique (Cf. Chapitre 8) sur la base d'une analyse question par question.

5.3 Les politiques connexes

Il s'agit d'identifier les principales interventions connexes et de situer l'intervention évaluée par rapport à elles.

L'identification des politiques connexes est utile pour :

- Aider à choisir les questions appartenant aux familles de la cohérence/complémentarité externe et de la pertinence (Cf. Section 6.4)
- Si nécessaire, pour examiner la qualité de la conception, et notamment le fait que les objectifs sont complémentaires de ceux des autres politiques connexes.

Le rapport d'évaluation expose succinctement les éléments suivants :

- Principaux organismes nationaux et internationaux intervenant dans le même territoire, dans le même secteur et/ou en direction du même public
- Principales politiques mises en œuvre par ces organismes dans le même territoire, dans le même secteur et/ou en direction du même public.

Certaines questions d'évaluation peuvent porter sur les politiques connexes pour permettre de travailler plus en profondeur, par exemple :

- L'intervention est-elle complémentaire et cohérente avec les politiques et les priorités du pays partenaire ? ... avec les autres politiques communautaires ?
- La mise en œuvre a-t-elle fait l'objet d'une coordination avec les actions des autres bailleurs de fonds et la complémentarité en a-t-elle été améliorée ?
- Dans quelle mesure y a-t-il eu une valeur ajoutée à conduire l'intervention au niveau de la Commission européenne plutôt qu'à celui des Etats membres ?

6 Questions d'évaluation

RÉSUMÉ

Utilité des questions

Les questions servent à focaliser le travail d'évaluation sur un petit nombre de points clés afin de permettre une collecte d'information plus ciblée, une analyse plus approfondie et un rapport plus utile.

Origine des questions

Les questions d'évaluation peuvent dériver directement ou indirectement de la logique d'intervention. Elles peuvent également porter sur des impacts inattendus, ne figurant pas dans la logique d'intervention, ou sur des aspects transversaux (genre, environnement, etc.).

Les questions

Les questions (une dizaine au maximum) sont sélectionnées selon l'utilité potentielle des réponses et leur faisabilité. Le choix des questions fait l'objet d'un débat au stade de la réunion de démarrage.

Questions et critères d'évaluation

Les questions se classent en différentes familles : les critères de l'OCDE - DAC (pertinence, efficacité, efficience, durabilité, impact) et deux critères propres à la Commission européenne (cohérence/complémentarité, valeur ajoutée communautaire).

Préparer une question d'évaluation

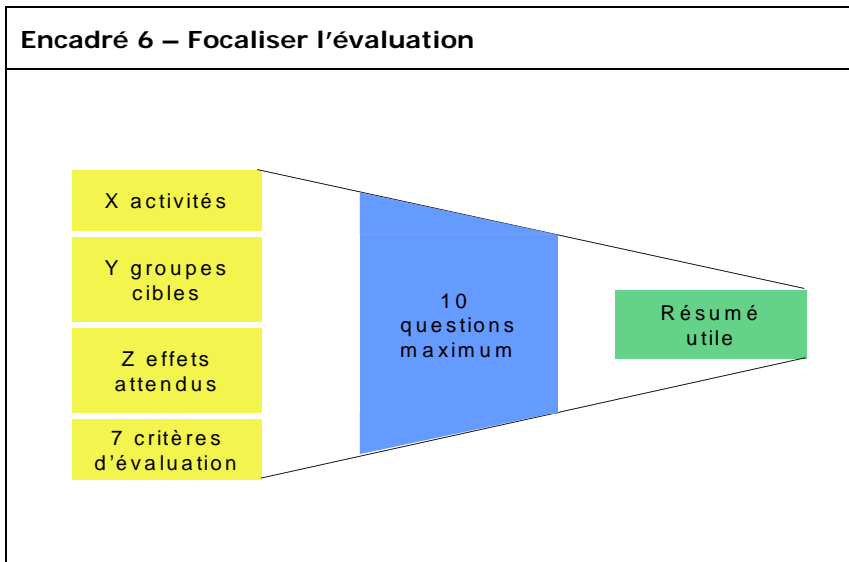
Il importe de vérifier que la réponse à la question sera utile, de préciser la nature de l'utilisation attendue, de vérifier que la question relève d'une évaluation, de préciser le champ de la question, de la relier à la logique d'intervention et à un critère d'évaluation.

6.1 Utilité des questions

Les questions servent à focaliser le travail d'évaluation sur un petit nombre de points clés afin de permettre une collecte d'information plus ciblée, une analyse plus approfondie et un rapport plus utile.

Focaliser une évaluation sur quelques questions est particulièrement nécessaire quand l'intervention évaluée est complexe (Cf. Encadré 1). Dans ce cas, si l'on voulait traiter toutes les dimensions de l'aide et toutes les dimensions de l'évaluation, le travail serait extrêmement long et coûteux ou bien alors extrêmement superficiel. Il est donc indispensable de faire des choix.

Comme le montre l'encadré suivant, le choix des questions contribue à optimiser l'utilité de l'évaluation.



6.2 Origine des questions

6.2.1 Questions déduites directement de la logique d'intervention

La logique d'intervention est l'ensemble des liens de causalité permettant d'expliquer comment l'intervention va atteindre ses objectifs et produire les effets attendus (Cf. Chapitre 5).

Lorsque l'équipe d'évaluation a identifié les effets attendus et les liens de causalité qui les relient, il devient possible de poser toutes sortes de questions du type :

- Dans quelle mesure [l'activité X] a-t-elle contribué à [générer l'effet Z] ?
- Dans quelle mesure [les activités X1, X2, X3...] ont-elles contribué à [générer l'effet Z] ?
- Dans quelle mesure [les activités X1, X2, X3...] ont-elles contribué à [générer les effets Z1, Z2, Z3, ...] ?

Ces questions dérivent directement de la logique d'intervention sous une forme standard et appartiennent à la famille de l'efficacité (obtention d'un ou plusieurs effets attendus).

6.2.2 Questions déduites indirectement de la logique d'intervention

Chaque question de forme standard peut être déclinée en de multiples autres questions,

... en précisant le champ, par exemple :

- Dans quelle mesure l'intervention [et plus spécifiquement celle mise en œuvre au moyen de l'instrument A ou de la procédure B] a-t-elle contribué à générer l'effet Z ?
- Dans quelle mesure [la coordination avec les autres partenaires du développement] a-t-elle contribué à générer l'effet Z ?

... ou en précisant l'effet concerné, par exemple :

- Dans quelle mesure l'intervention a-t-elle contribué à générer l'effet Z [pour les catégories les plus pauvres de la population] ?

... ou en changeant le critère d'évaluation (Cf. Section 6.4), par exemple :

- Dans quelle mesure l'intervention a-t-elle généré l'effet Z [avec une forte probabilité de survie des effets après la fin de l'aide] ? (durabilité)
- Dans quelle mesure l'intervention a-t-elle contribué à générer l'effet Z [avec un coût limité comparativement à ...] ? (efficacité ou coût-efficacité)
- Lorsque l'intervention vise à obtenir l'effet Z, dans quelle mesure cela [correspond-il aux besoins de la population concernée] ? (pertinence)
- Lorsque l'intervention vise à obtenir l'effet Z, dans quelle mesure cela [est-il compatible ou contradictoire avec les

objectifs des autres politiques de la Commission européenne] ? (cohérence/complémentarité)

- Lorsque l'intervention vise à obtenir l'effet Z, dans quelle mesure cela [apporte-t-il une valeur ajoutée par rapport à une intervention similaire qui aurait été conduite par les Etats membres] ? (valeur ajoutée communautaire).

6.2.3 Questions non déduites de la logique d'intervention

Les questions qui suivent ne nécessitent pas un examen préalable de la logique d'intervention parce qu'elles concernent des effets qui n'y figurent pas :

- Dans quelle mesure [l'activité X, l'instrument A, la procédure B] a-t-elle généré des effets inattendus ? Et si oui au bénéfice ou au détriment de qui ?
- Dans quelle mesure la Commission européenne a-t-elle intégré [l'aspect transversal T] dans la conception et la mise en œuvre de ses interventions ?

6.3 Choisir les questions

Le choix des questions d'évaluation se fait en plusieurs étapes :

Faire une première liste

Une première version des questions d'évaluation est proposée sur la base :

- De l'analyse de la logique d'intervention
- De l'analyse de la raison d'être de l'intervention
- Des sujets qui ont justifié la décision de lancer l'évaluation
- Des sujets à étudier, comme précisé dans les termes de référence
- Des questions posées dans l'évaluation *ex ante*, le cas échéant.

Ensuite, la liste et la formulation des questions prennent en compte également :

- Les sujets soulevés par les informateurs clés au début de l'évaluation
- Les attentes des membres du groupe de référence

- Les sujets découverts par l'évaluateur

Apprécier l'utilité potentielle des réponses

Dans un second temps, il convient d'apprécier l'utilité potentielle des réponses, en faisant l'hypothèse que l'évaluation apportera des réponses de bonne qualité. Pour chaque question envisagée, il s'agit de savoir :

- Qui devrait utiliser la réponse ?
- Quelle est l'utilisation attendue : connaissance, négociation, décision, communication ?
- La réponse n'est-elle pas déjà connue ?
- N'y a-t-il pas une autre étude (audit, bilan, revue, etc...) en cours de réalisation et susceptible d'apporter des réponses ?

Si le choix des questions doit faire l'objet d'une discussion en réunion, il peut être intéressant de les classer en trois catégories d'utilité potentielle : plus élevée, moyenne, moins élevée.

Vérifier qu'il n'y a pas eu d'oubli grave

L'expérience montre que les oublis les plus fréquents et donc préjudiciables à la qualité d'une évaluation concernent :

- Les questions d'efficience et de durabilité
- Les questions portant sur les effets négatifs, particulièrement si ces effets concernent des groupes défavorisés
- Les questions concernant les effets à très long terme.

Apprécier la faisabilité des questions

La faisabilité (évaluabilité) d'une question doit être examinée mais toujours après son utilité. Dans ce but, il convient de consulter :

- Le service gestionnaire de l'intervention
- Un ou plusieurs experts du domaine
- Un ou plusieurs professionnels de l'évaluation.

Limites les plus fréquentes à l'évaluabilité

- **Des faiblesses de gestion**

Les données et rapports de suivi sont inadéquats ou peu fiables.

Les difficultés de gestion ont généré des conflits qui limitent l'accès à certains informateurs ou biaisent leurs réponses.

Dans le cas d'une forte suspicion de pratiques illicites ou illégales, il est préférable de reporter l'évaluation et de lancer un audit.

- **Un champ trop complexe**

Au regard du temps et du budget disponibles, il y a beaucoup trop de données à collecter, de personnes à rencontrer, et d'analyses à effectuer.

- **Une activité trop récente**

L'activité n'a pas encore produit d'effets observables.

Les personnes interrogées n'ont pas encore stabilisé leurs opinions.

- **Une activité trop innovante**

Il est difficile de définir les termes de la question sans ambiguïté.

L'expertise est insuffisante pour comprendre les mécanismes de causalité.

- **Une intervention trop marginale**

L'activité évaluée n'atteint pas la masse critique qui permettrait une analyse de sa contribution.

- **Un impact trop lointain**

Il y a beaucoup de facteurs externes et ils sont tellement influents qu'il devient impossible d'analyser la contribution de l'intervention.

Si une question est potentiellement très utile mais difficile à traiter, il convient de rechercher si une question voisine serait à la fois plus facile et également utile. Par exemple, si une question porte sur un impact relativement lointain ou global, on pourra probablement améliorer sa faisabilité en la faisant porter sur un ou

quelques impacts immédiatement précédents dans la logique d'intervention.

Discuter du choix des questions

Le choix des questions fait l'objet d'un débat avec le groupe de référence au stade de la réunion de démarrage.

La sélection est d'autant plus réussie que les utilisateurs potentiels ont été consultés, se sont approprié les questions sélectionnées et qu'aucun point de vue légitime n'a été écarté.

Raisons pour sélectionner une question

- **Parce que quelqu'un l'a posée**

Celui qui demande qu'une question soit traitée tend à coopérer à la construction de la réponse et à s'approprier les conclusions.

Il arrive qu'un acteur pose une question avec l'intention principale d'influencer ou même de contrarier l'action d'un autre acteur. L'utilité potentielle de ce type de question doit être examinée avec attention.

- **Parce qu'elle est utile, par exemple :**

L'intervention ou un de ses aspects est innovant et plusieurs acteurs attendent une validation.

Une décision va être prise et les conclusions peuvent arriver à temps pour aider à prendre cette décision.

Un débat public est prévu et les conclusions peuvent arriver à temps pour alimenter ce débat.

- **Parce la réponse n'est pas connue d'avance, on évite donc de poser une question si :**

Une autre évaluation, un audit ou une étude vient d'y répondre.

Elle a déjà été posée dans de nombreuses autres évaluations avec toujours la même réponse.

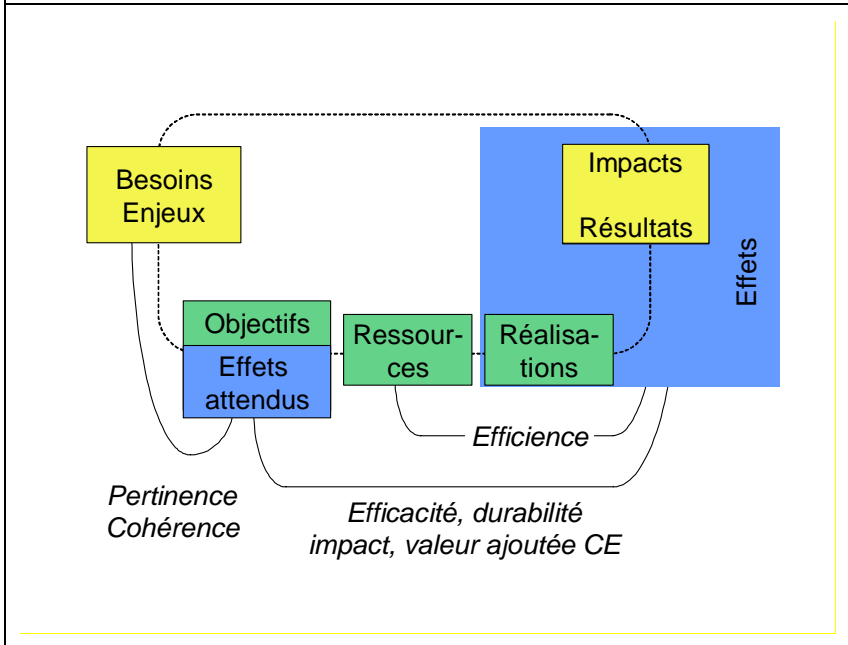
Sur le site Web : exemples de questions pour une évaluation au niveau pays, au niveau région, pour une évaluation thématique ou sectorielle

6.4 Questions et critères d'évaluation

Les questions se classent en différentes familles qui correspondent à différents « angles de vue » sur ce qui est évalué. Sept de ces points de vue, appelés également critères d'évaluation, doivent être considérés : pertinence, efficacité, efficience, durabilité, impact, cohérence/ complémentarité, et valeur ajoutée communautaire.

Les cinq premiers correspondent à la pratique traditionnelle de l'évaluation de l'aide au développement formalisée par l'OCDE (CAD). Les deux suivants s'appliquent à l'ensemble des politiques de la Commission européenne.

Encadré 7 – Principaux critères d'évaluation



Pertinence

Dans quelle mesure les objectifs de l'action de développement correspondent-ils aux attentes des bénéficiaires, aux besoins du pays, aux priorités globales, et aux politiques des partenaires et des bailleurs de fonds ?

- Exemple : Dans quelle mesure la concentration de l'aide sur l'éducation de base correspond-elle aux besoins du pays partenaire ?

Efficacité

Dans quelle mesure les objectifs de l'action de développement ont-ils été atteints, ou sont-ils en train de l'être, compte tenu de leur importance relative ?

- Exemple : Dans quelle mesure l'aide a-t-elle contribué à un accès équitable à une éducation de base de bonne qualité ?

Efficience

Dans quelle mesure les effets désirés sont-ils obtenus avec le moins de ressources possibles (fonds, expertise, temps, coûts administratifs) ?

- Exemple : La mise en œuvre sous forme d'aide budgétaire sectorielle a-t-elle permis d'obtenir des effets comparables à ceux d'une approche par projet avec un coût de coordination moindre pour la Commission européenne et le pays partenaire ?

Durabilité (pérennité, viabilité)

Dans quelle mesure les bénéfices résultant de l'action de développement continuent-ils après la fin de l'intervention de la Commission européenne, ou vont-ils probablement continuer sur le long terme en résistant aux risques ?

- Exemple : Dans quelle mesure l'aide a-t-elle contribué à un rattrapage durable du retard d'entretien du réseau routier ?

Impact

Effets à long terme, positifs et négatifs, primaires et secondaires, induits par une action de développement, directement ou non, intentionnellement ou non.

- Exemple : Du point de vue des publics concernés, les nuisances environnementales sont-elles acceptables en comparaison des effets positifs de l'intervention ?

Cohérence/complémentarité

Ce critère peut avoir plusieurs dimensions :

Cohérence à l'intérieur du programme de développement de la Commission.

- Exemple : Peut-on dire que les activités et les réalisations permettent logiquement d'atteindre les objectifs ? Y a-t-il des contradictions entre les différents niveaux d'objectifs ? Y a-t-il des doubles emplois entre les activités ?

Cohérence/complémentarité avec les politiques du pays partenaire et complémentarité avec les interventions des autres bailleurs de fonds.

- Exemple : Peut-on dire qu'il n'y a pas de double emploi entre l'intervention considérée et d'autres interventions du pays partenaire et/ou d'autres bailleurs de fonds et en particulier des Etats membres ?

Cohérence/complémentarité avec les autres politiques communautaires.

- Exemple : Y a-t-il convergence entre les objectifs de l'action de développement et ceux des autres politiques communautaires (commerce, agriculture, pêche, etc.) ?

Valeur ajoutée communautaire

Dans quelle mesure les avantages apportés par l'action de développement s'ajoutent-ils à ceux qui auraient résulté de la seule intervention des Etats membres dans le pays partenaire ?

- Exemple : Dans quelle mesure le partage des rôles entre la Commission européenne et les Etats membres a-t-il contribué à optimiser l'impact de l'aide ?

6.5 Préparer une question d'évaluation

6.5.1 Vérifier que la question relève d'une évaluation

Avant de travailler à la rédaction d'une question, on doit s'assurer qu'elle ne concerne pas l'audit ou le suivi.

Si la question traite uniquement du respect des règles ou des procédures, elle relève d'un audit et non pas d'une évaluation.

Si la question traite uniquement de l'avancement des réalisations, elle relève du suivi.

6.5.2 Préciser le champ de la question

Le champ de la question peut être l'intervention toute entière ou une dimension plus particulière de sa conception ou de sa mise en œuvre, par exemple :

- L'intervention a-t-elle contribué à générer l'effet Z comme attendu ? (ensemble de l'intervention)
- La participation des acteurs à la conception de l'intervention a-t-elle permis de mieux générer l'effet Z ? (une modalité de conception)
- Les mesures prises pour assurer la coordination avec les autres bailleurs ont-elles permis de mieux générer l'effet Z ? (une modalité de mise en œuvre)
- Le choix d'un appui budgétaire sectoriel a-t-il permis de mieux générer l'effet Z ? (une modalité de financement).

6.5.3 Relier la question à la logique d'intervention

Sur quel effet la question porte-t-elle ? Par exemple :

- L'intervention a-t-elle contribué à générer l'effet Z comme attendu ?
- L'intervention a-t-elle contribué à générer l'effet Z pour les publics les plus pauvres ?
- L'intervention a-t-elle augmenté les chances des publics les plus pauvres d'obtenir l'effet Z ?

Si la question porte sur un résultat à court terme ou un impact spécifique, elle sera plus facilement évaluable que si elle porte sur un impact global. A l'inverse, une question portant sur un impact plus global aura une plus grande utilité pour les décideurs au niveau stratégique.

6.5.4 Rattacher la question à un critère d'évaluation

La question est rédigée en ayant à l'esprit un critère d'évaluation (ex : pertinence, efficacité, efficience, durabilité, impact, valeur ajoutée communautaire et cohérence/complémentarité,

(Cf. Section 6.4), sauf si elle vise seulement à connaître ou comprendre.

6.5.5 Finaliser la rédaction de la question

Il est préférable de choisir une rédaction ouverte appelant une réponse nuancée plutôt qu'une question fermée appelant une réponse de forme « oui/non ». Par ailleurs, la rédaction de la question doit être simple et concise. Enfin, il est judicieux d'ajouter à titre indicatif des commentaires sur tout ou partie des points suivants :

- Précisions sur le champ de la question
- Définition des principaux termes utilisés.

Sur le Web : exemples de questions pour différentes catégories d'évaluation

7 Référentiel de jugement

RÉSUMÉ

Critères de jugement

Un critère de jugement (également appelé critère d'appréciation raisonnée) précise un aspect des mérites ou du succès de l'intervention évaluée. Il permet de répondre de façon positive ou négative à une question d'évaluation.

Niveau cible

Pour formuler un jugement explicite, il est souhaitable de faire référence à la fois à un critère de jugement et à un niveau cible, c'est-à-dire à une norme de performance, un seuil de succès ou une bonne pratique qui sert de point de comparaison.

Indicateur

Les indicateurs précisent à l'avance quelles informations vont être collectées. Ils aident ainsi à concentrer le processus de collecte.

L'évaluation s'intéresse principalement aux indicateurs de résultats et d'impacts renseignés à l'aide d'enquêtes, de bases de données de gestion ou de sources statistiques.

De la question au critère et à l'indicateur

Ce chapitre propose une démarche pour passer des questions d'évaluation aux critères de jugement et aux indicateurs.

7.1 Critères de jugement

Un critère de jugement précise un aspect de l'intervention évaluée qui va permettre d'apprécier son mérite ou son succès. Bien que « critère de jugement » soit le terme approprié, une alternative acceptable est « critère d'appréciation raisonnée ».

Ne pas confondre les concepts !

Les critères de jugement présentés dans ce chapitre.

Les critères d'évaluation : pertinence, efficacité, efficience, durabilité, impact, valeur ajoutée communautaire et cohérence/complémentarité (Cf. Section 6.4).

Les critères de qualité des rapports d'évaluation (Cf. Section 12.1).

Un exemple de question est :

- Dans quelle mesure l'aide de la Commission européenne a-t-elle amélioré la capacité du système d'enseignement primaire à accueillir les élèves des groupes défavorisés sans discrimination ?

Cette question appartient à la famille de l'efficacité.

Comme la plupart des questions d'évaluation, celle-ci comporte deux parties :

- Ce qui est jugé : « l'aide de la Commission européenne »
- La façon de juger : « ... a-t-elle amélioré la capacité du système d'enseignement primaire à accueillir les élèves des groupes défavorisés sans discrimination ? »

Les critères de jugement développent et précisent la deuxième partie de la question, par exemple :

- Critère 1 : Capacité du système d'enseignement primaire à accueillir les élèves de la minorité ethnique E avec une qualité d'enseignement suffisante.
- Critère 2 : Capacité du système d'enseignement primaire à accueillir les élèves de zones urbaines défavorisées avec une qualité d'enseignement suffisante.

Pour être utilisé en pratique, chaque critère de jugement doit être accompagné d'un niveau cible (*target*) et d'un ou plusieurs indicateurs.

Recommandations

Toujours définir le critère de jugement avant de sélectionner un indicateur existant ou de concevoir un nouvel indicateur. Cela est indispensable pour clarifier les concepts. Si l'on porte trop vite attention aux indicateurs, on risque de rester prisonnier de l'information existante, même si elle est peu adaptée pour répondre à la question posée.

Discuter des critères de jugement avec le groupe de référence de façon à tenir compte de la diversité des points de vue pertinents à propos de l'intervention.

Il peut exister des désaccords sur les critères de jugement, par exemple un même effet peut avoir une dimension jugée positivement par certains membres du groupe de référence et une autre dimension jugée négativement par d'autres. Deux options sont possibles dans ce cas : (1) choisir un seul critère de jugement, mais il faut alors éviter de biaiser l'évaluation, ou (2) choisir plusieurs critères, mais la conséquence sera d'augmenter et de compliquer le travail de collecte et d'analyse d'information.

Pour optimiser la collecte et l'analyse de l'information, l'idéal est de définir un nombre limité de critères de jugement pour chaque question. Cette recommandation tient compte de la capacité des utilisateurs à s'approprier l'information.

Justifier, le cas échéant, les écarts qui pourraient exister entre les critères utilisés pour formuler le jugement à la fin du processus d'évaluation et ceux identifiés *a priori* dans la phase de documentation (*desk*).

7.2 Niveau cible (*target*)

Le concept de « niveau cible » est largement utilisé dans le contexte du management public pour établir un objectif vérifiable ou un niveau de performance à atteindre. Dans le contexte de l'évaluation, il est utilisé avec un sens encore plus large. En effet, le niveau cible peut être établi à l'avance, mais aussi identifié de façon *ad hoc* au moment de l'évaluation.

7.2.1 Objectif défini de façon vérifiable

Le niveau cible peut apparaître dans les objectifs de l'intervention, si toutefois ils ont été établis de façon vérifiable. Dans ce cas particulier, le même indicateur sert à définir l'objectif, à rendre le critère de jugement opérationnel et à déterminer le niveau cible.

- Exemple : le nombre d'enseignants qualifiés et expérimentés pour 1 000 enfants en âge d'aller en école primaire est au moins de 20.

7.2.2 Bonne pratique comparable

Le niveau cible est établi au cours de l'évaluation par référence à une bonne pratique comparable à l'intervention évaluée.

- Exemple : l'accès à une éducation primaire avec des professeurs qualifiés et expérimentés est au moins aussi satisfaisant que dans le cas de W (bonne pratique reconnue au niveau régional).

La procédure se déroule comme suit :

- Identifier une pratique comparable et reconnue pour sa qualité (intervention similaire de la Commission européenne dans un autre pays, intervention d'un autre bailleur, intervention dans un autre secteur mais qui utilise les mêmes instruments)
- Se renseigner sur la pratique à comparer (cela est plus facile si elle a déjà été évaluée)
- Vérifier que les conditions de contexte sont suffisamment proches pour permettre la comparaison
- Procéder à la comparaison (essentiellement qualitative)
- Discuter et valider la comparaison avec le groupe de référence.

7.2.3 Meilleures pratiques identifiées au sein de l'intervention

Le niveau cible est établi à la fin de l'évaluation par référence à une bonne pratique qui a été identifiée au sein de l'intervention évaluée.

- Exemple : dans des zones où la minorité ethnique X est concentrée, le nombre d'enseignants qualifiés et expérimentés pour 1000 enfants en âge d'aller à l'école pri-

maire est comparable à celui de la zone Z, considérée comme une des plus performantes parmi les autres zones rurales du pays.

7.3 Indicateurs

Le recours aux indicateurs est nécessaire pour :

- Collecter et traiter les données sous une forme qui sera directement utilisable au moment de répondre aux questions
- Aller à l'essentiel dans la collecte d'information.

Un indicateur quantitatif est fondé sur un comptage (ex : nombre d'enseignants qualifiés et expérimentés). L'indicateur élémentaire est le comptage proprement dit. Il peut être utilisé pour calculer des indicateurs dérivés (ratios, taux) tels que le coût par élève scolarisé ou le nombre d'enseignants qualifiés et expérimentés pour 1 000 enfants en âge de fréquenter l'école primaire.

Un indicateur qualitatif prend la forme d'une affirmation qui devra être vérifiée lors de la collecte d'informations (ex : l'opinion des parents est que leurs enfants ont la possibilité d'être admis dans une classe d'enseignement primaire avec un professeur qualifié et expérimenté).

Attention !

Des données non structurées sont également collectées durant l'évaluation, soit de façon incidente, soit parce qu'on a employé des outils tels que des études de cas. Il s'agit d'éléments de preuve qui peuvent avoir une force suffisante pour fonder des conclusions, mais ce ne sont pas des indicateurs.

7.3.1 Indicateurs d'évaluation

Les principaux indicateurs d'évaluation sont ceux qui sont liés aux critères de jugement et qui précisent quelles informations doivent être collectées pour porter un jugement selon ces critères.

Un indicateur peut être construit spécifiquement pour une évaluation (indicateur *ad hoc*) et renseigné par exemple dans le cadre d'une enquête. Il peut aussi être tiré des bases de données

de suivi, d'un système d'appréciation de performance ou de sources statistiques.

Les indicateurs d'évaluation et les autres

Quand une question d'évaluation concerne un résultat ou un impact attendu, il convient de vérifier si ce résultat ou cet impact a fait l'objet d'un suivi de performance. Dans ce cas, l'équipe d'évaluation utilise les indicateurs et les données correspondantes, ce qui constitue une aide considérable, surtout si le point zéro a été enregistré.

Le suivi des performances peut cependant se révéler peu ou pas utile dans le cas où une question d'évaluation concerne les aspects transversaux, les facteurs de durabilité, les effets non attendus, l'évolution des besoins ou des problèmes, la cohérence, etc.

7.3.2 Catégories d'indicateurs

Indicateurs et cycle d'intervention

Des indicateurs sont utilisés tout d'abord pour l'analyse du contexte et la formulation de la stratégie d'intervention, ensuite pour le suivi des réalisations et des résultats et enfin pour l'évaluation.

Indicateurs de suivi et de mesure de performance

Les systèmes de suivi et de mesure de performance utilisent des indicateurs dérivés du diagramme des effets attendus (également appelés chaînes de résultats).

Les indicateurs de suivi sont surtout liés aux ressources et réalisations. Les indicateurs de performance sont surtout concentrés sur les résultats.

Indicateurs et évaluation

Les indicateurs d'évaluation sont utilisés pour aider à répondre aux questions d'évaluation spécifiques. Suivant la question, ils peuvent être liés aux besoins, problèmes et défis qui ont justifié l'intervention, ou à l'obtention des réalisations, résultats et impacts attendus.

Indicateurs globaux et spécifiques

Les indicateurs globaux ou contextuels s'appliquent à l'ensemble d'un territoire, d'une population ou d'une catégorie de population, sans faire de distinction entre ceux qui ont été atteints par l'intervention et ceux qui ne l'ont pas été. Ils sont principalement recherchés dans les bases de données statistiques. Leur définition est établie en dehors de l'évaluation. Un chapitre du Volume 4 y est consacré.

Les indicateurs spécifiques concernent seulement la partie ou la catégorie du public ou la partie d'un territoire qui a été effectivement atteinte. Les indicateurs spécifiques permettent de suivre à la trace les changements qui apparaissent chez ceux qui ont été atteints par l'intervention. Ils sont quelquefois définis et testés par l'équipe d'évaluation, par exemple sous la forme d'une échelle de notation *ad hoc*. En général, ils sont renseignés au moyen d'enquêtes ou à partir des bases de données de gestion.

Sur le site Web : comment construire une grille de notation

Indicateurs et logique d'intervention

Les indicateurs peuvent servir à produire une information sur les ressources, les réalisations, les résultats ou les impacts.

Encadré 8 – Indicateurs et logique d'intervention

Catégorie	... informent sur :
Indicateurs de ressources	Les moyens financiers, humains, matériels, organisationnels ou réglementaires mobilisés pour la mise en œuvre de l'intervention.
Indicateurs d'activités	Les processus de mise en œuvre et la gestion.
Indicateurs de réalisations	Les produits et services délivrés sous la responsabilité des gestionnaires de l'intervention. Pour simplifier, on peut dire que les réalisations correspondent à ce qui est acheté avec l'argent public.
Indicateurs de résultats	Les effets immédiats de l'intervention pour ses destinataires directs. Un effet est immédiat si l'opérateur peut le connaître facilement pendant qu'il est au contact du destinataire. Étant facilement connus des opérateurs, les indicateurs de résultats directs peuvent souvent être quantifiés de façon exhaustive par le système de suivi.
Indicateurs d'impacts	Les conséquences plus lointaines et/ou plus indirectes de l'intervention.

7.4 De la question à l'indicateur

L'encadré 9 montre un exemple de passage de la question à l'indicateur. Il faut garder à l'esprit qu'une question est généralement déclinée en plusieurs critères de jugement, eux-mêmes déclinés en plusieurs indicateurs.

Dans l'exemple suivant, la question fait référence à une famille de critères d'évaluation : l'efficacité.

Le critère de jugement est dérivé de la question de la façon suivante :

- Il montre le type de succès que l'on cherche à évaluer en posant la question.
- Il précise plusieurs concepts tels que « système éducatif », « groupes défavorisés » et « discrimination ».

Encadré 9 – De la question à l'indicateur : exemple

Question	Dans quelle mesure l'aide de la Commission européenne a-t-elle amélioré la capacité du système éducatif à accueillir les élèves des groupes défavorisés sans discrimination ?
Critère(s) de jugement	Capacité du système d'enseignement primaire à accueillir les élèves de la minorité ethnique E avec une qualité satisfaisante.
Indicateur(s)	Nombre d'enseignants qualifiés et expérimentés pour 1000 enfants en âge de fréquenter l'enseignement primaire dans les zones où la minorité ethnique E est concentrée.

L'indicateur est dérivé du critère de jugement de la façon suivante :

- Il décrit de façon précise l'information qui doit être recherchée pour juger selon le critère de jugement choisi.
- L'indicateur est fidèle au critère de jugement (« capacité du système d'enseignement primaire à accueillir les élèves avec une qualité satisfaisante »).
- Il est quantitatif, mais on aurait pu définir un indicateur qualitatif, par exemple : « les parents interrogés confirment que leurs enfants ont la possibilité de fréquenter une classe primaire bénéficiant d'un enseignant qualifié et expérimenté ».

7.4.1 Qualité d'un indicateur

Un indicateur mesure ou qualifie précisément le critère de jugement ou la variable que l'on cherche à observer. À défaut on peut avoir recours à plusieurs indicateurs approchés (*proxies*).

L'indicateur fournit une information simple, facilement communicable et comprise de la même façon par le fournisseur et l'utilisateur de l'information. Il est précis, c'est-à-dire qu'il est associé à une définition qui ne comporte aucun terme ambigu. Il est sensible, c'est-à-dire qu'il génère des données qui varient significativement s'il y a un changement dans ce que l'on observe.

Il est souvent demandé aux indicateurs de performance et aux niveaux cibles d'être SMART (Spécifiques, Mesurables, Ambitieux, Réalistes, et produits dans les Temps). La qualité d'un indicateur d'évaluation s'apprécie différemment. Le plus important est que l'indicateur soit fidèle au critère de jugement.

Indicateurs et effets : Attention !

L'indicateur employé pour l'évaluation d'un effet n'est pas en lui-même une mesure ou une preuve de cet effet. L'indicateur informe seulement sur les changements intervenus, lesquels peuvent résulter de l'intervention (effet) ou d'autres causes.

Il est toujours nécessaire que l'équipe d'évaluation analyse ou interprète l'indicateur pour estimer un effet (Cf. Chapitre 10).

8 Conception méthodologique

RÉSUMÉ

Il s'agit de mettre en place l'approche qui permettra à l'équipe d'évaluation de répondre aux questions et d'arriver à un jugement d'ensemble. La conception méthodologique de l'évaluation inclut : une stratégie de collecte et d'analyse des données, des terrains sélectionnés pour les investigations, une série d'outils développés spécifiquement et un plan de travail.

Tableau de conception par question

Le tableau de conception explique comment on va répondre à une question d'évaluation, y compris la chaîne de raisonnement qui va relier les données, les constats et les conclusions. Il décline : le champ de la question, les critères de jugement, les indicateurs, les niveaux cibles, la chaîne de raisonnement logique, la stratégie d'analyse, les terrains d'investigation et les sources d'information et outils.

Chaîne de raisonnement

La chaîne de raisonnement logique montre comment l'équipe d'évaluation projette de répondre à la question. Il ne s'agit pas de complexifier la question mais d'aider à y répondre.

Optimisation de la méthode

Il s'agit de finaliser la méthode d'évaluation dans son ensemble de façon à couvrir l'ensemble des questions évaluatives et à disposer d'une bonne combinaison d'outils d'évaluation, en considérant le temps disponible et les ressources.

8.1 Introduction

La conception méthodologique consiste à mettre en place la démarche qui permettra à l'équipe d'évaluation de répondre aux questions et d'arriver à un jugement d'ensemble. Elle inclut :

- La chaîne de raisonnement logique qui sera suivie
- La stratégie de collecte et d'analyse des données
- Les terrains sélectionnés pour les investigations
- Une série d'outils développés spécifiquement

- Le plan de travail.

L'équipe d'évaluation commence à concevoir une méthode provisoire dès la rédaction de sa proposition afin d'estimer les coûts. Une hypothèse clé à ce stade consiste à déterminer jusqu'à quel point l'évaluation pourra s'appuyer sur des données secondaires ou nécessitera des travaux spécifiques de collecte d'information de terrain.

Les grandes lignes de la méthode sont tracées pendant l'étape de démarrage, en fonction des questions d'évaluation, des critères de jugement, des indicateurs, des outils de collecte de données et de la stratégie d'analyse.

La méthode est affinée et finalisée avant la phase de terrain et complètement décrite dans le rapport de documentation (*desk*).

Le rapport final inclut une présentation courte et pointue de la méthode d'évaluation, ainsi que ses limites s'il y en a. La méthode est complètement décrite dans une annexe, y compris la conception initiale, les problèmes rencontrés, les solutions trouvées, la méthode réellement appliquée, et les limites.

La méthode d'évaluation est conçue selon un processus itératif à trois niveaux :

- Une approche question par question, permettant à l'équipe d'évaluation de préparer les tableaux de conception en vue d'apporter une réponse adéquate à chacune des questions
- Une approche d'ensemble qui couvre toutes les questions, et qui permet à l'équipe d'évaluation d'optimiser sa méthode dans son ensemble, tout en respectant les contraintes de temps et de ressources
- Un ensemble d'outils développés pour collecter et analyser l'information.

8.2 Tableau de conception par question

Le tableau de conception explique comment on va répondre à une question d'évaluation, y compris la chaîne de raisonnement qui va relier les données, les constats et les conclusions.

Un tableau de conception est élaboré pour chaque question.

Encadré 10 – Structure d'un tableau de conception

Question	Texte de la question
Commentaire	Pourquoi cette question est-elle posée ?
Champ	Qu'est ce que la question couvre ?
Critère de jugement	Comment les mérites ou les succès seront-ils appréciés ?
Indicateur(s)	Quelles sont les données qui aideront à apprécier les mérites ou les succès ?
Niveau(x) cible(s)	Quel niveau ou seuil peut être considéré comme un succès ?
Chaîne de raisonnement	Étapes de raisonnement prévues pour répondre à la question en : - Renseignant les indicateurs - Analysant l'information - Formulant le jugement
Stratégie d'analyse	Type(s) d'analyse à entreprendre (Cf. Section 10.1.1)
Terrains d'investigation	Niveau auquel les données seront collectées et traitées (Cf. Section 8.3.2)
Sources d'information et outils	D'où viendront les informations (Cf. Section 8.3.3)

Sur le site Web : exemple de tableau de conception

Ce tableau est progressivement affiné en versions successives :

- Version préliminaire annexée au rapport de démarrage
- Versions provisoires préparées pendant la phase documentaire de l'évaluation jusqu'à ce que la conception méthodologique soit progressivement optimisée

- Version finale annexée au rapport de la phase de documentation (*desk*).

8.2.1 Chaîne de raisonnement

La chaîne de raisonnement décrit toutes les étapes qui vont se succéder pour répondre à la question. Il ne s'agit en aucun cas de complexifier la question. Le raisonnement prévu est indicatif mais il vaut mieux le clarifier à l'avance car :

- Les membres du groupe de référence ont ainsi l'occasion d'apporter des avis et des contributions spécifiques
- Tous les membres de l'équipe d'évaluation comprennent pourquoi ils collectent et traitent les données et ils travaillent donc de manière plus efficace, homogène et cohérente
- Les membres de l'équipe qui ne sont pas familiers de l'évaluation reçoivent des conseils utiles sur les données qui doivent être collectées et traitées.

Étapes concernant les indicateurs

Elles peuvent concerner :

- (1) le niveau ou état actuel des indicateurs, éventuellement décomposé par pays, zone, groupe social, etc., par exemple :
 - Quelle est la valeur actuelle de l'indicateur quantitatif I au niveau national, et pour les zones/groupes ciblés ?
- (2) les changements dans les indicateurs, par exemple :
 - Les parties prenantes ont-elles perçu un changement positif dans l'indicateur qualitatif J, au cours de la période évaluée ?

Étapes concernant l'analyse

Elles sont écrites en vue de :

- (3) Confirmer les hypothèses sur le succès de l'intervention et justifier une réponse positive à la question, par exemple :
 - La conception de l'aide de la Commission européenne inclut-elle un engagement à suivre des indicateurs de performance relatifs à l'effet Z ?
 - Un tel suivi a-t-il été réalisé ?

- Les données de suivi ont-elles fait l'objet de discussions périodiques avec les partenaires ?
 - Les partenaires ont-ils pris des mesures à la suite de telles discussions ?
 - Les mesures prises visaient-elles à obtenir l'effet Z ?
 - Etc.
- (4) Contester les hypothèses sur le succès de l'intervention et justifier une réponse négative à la question, par exemple :
- D'autres partenaires ont-ils également fait la promotion d'un suivi des indicateurs relatifs à l'effet Z ?
 - Les acteurs non-gouvernementaux ont-ils contribué à mettre en avant la question de l'effet Z dans le débat public ?
 - Jusqu'à quel point les partenaires ont-ils influencé les actions en faveur des groupes désavantagés ?

Étapes concernant l'application de critères de jugement

Elles sont écrites en vue de :

- (5) appliquer et affiner les critères de jugement dans le contexte spécifique de l'intervention, par exemple :
- Les parties prenantes pensent-elles spontanément aux mêmes critères de jugement que ceux choisis pour l'évaluation ? Si non, pourquoi ?
- (6) appliquer ou développer des niveaux cibles dans le contexte spécifique de l'intervention, par exemple :
- Quels sont les territoires/groupes qui ont les meilleures performances du pays en ce qui concerne les critères de jugement choisis ? Parmi eux, lesquels peuvent être légitimement comparés aux territoires/groupes ciblés ?

8.2.2 Terrains d'investigation

L'équipe d'évaluation peut vouloir collecter et traiter les données au niveau de l'intervention dans son ensemble, ou étudier plus spécifiquement quelques terrains, par exemple :

- Toutes les étapes du raisonnement seront parcourues dans le cadre d'une sélection de pays et s'accompagneront d'une visite dans chaque pays, ainsi que d'une note par pays.

- En plus de l'utilisation des statistiques nationales, l'équipe d'évaluation va étudier une sélection de districts typiques respectivement (a) du groupe/territoire ciblé et (b) des groupes/territoires les plus performants.
- L'équipe d'évaluation choisira soigneusement N projets qui feront l'objet d'une recherche détaillée afin de traiter certaines étapes du raisonnement.

8.2.3 Sources d'information et outils

Le tableau de conception identifie les sources d'information à utiliser, comme :

- Des statistiques comprenant des indicateurs de contexte disponibles dans les bases de données internationales (Cf. Volume 4)
- Des bases de données de gestion ou de suivi
- Des rapports, audits, évaluations, articles, etc.
- Des déclarations de parties prenantes
- Des déclarations d'experts.

Dans la plupart des cas, une source d'information est associée à un outil d'évaluation, par exemple une catégorie donnée de parties prenantes sera atteinte à travers des entretiens ou des *focus groups*, l'opinion des experts sera obtenue par des échanges de courriels ou par un panel, etc.

8.3 Optimisation de la conception d'ensemble

Il s'agit de finaliser la méthode d'évaluation dans son ensemble de façon à couvrir toutes les questions d'évaluation et à disposer d'une bonne combinaison d'outils, en considérant le temps disponible et les ressources.

Plusieurs itérations peuvent être nécessaires pour permettre à l'équipe d'évaluation d'optimiser la conception d'ensemble tout en assurant une qualité élevée pour chaque question.

Les itérations successives méritent réflexion mais ne prennent pas beaucoup de temps et se réalisent essentiellement au cours de la phase de démarrage (*inception*) et de documentation (*desk*).

8.3.1 Combiner outils et questions

L'équipe d'évaluation dresse la liste des outils d'évaluation envisagés dans les tableaux de conception. Chaque outil est alors considéré du point de vue de sa capacité à aider à répondre à plusieurs questions.

Elle explique et justifie ses principaux choix techniques, ainsi que les options alternatives, leurs avantages et inconvénients, et les risques associés.

L'équipe d'évaluation apprécie si les outils choisis peuvent se renforcer mutuellement, par exemple :

- Une série d'entretiens peut aider à identifier les documents pertinents en vue d'une analyse documentaire
- Une série d'entretiens peut aider à préparer un *focus group*
- Une série d'entretiens peut aider à affiner un questionnaire à passer à un échantillon de bénéficiaires
- Une série de monographies d'étude de cas peut être soumise à l'examen d'un panel d'experts en vue de renforcer l'analyse et d'aboutir à des constats plus précis.

8.3.2 Préparer le jugement d'ensemble

L'équipe d'évaluation porte un regard transversal sur tous les tableaux de conception en vue de préparer sa synthèse finale, c'est-à-dire un jugement d'ensemble fondé sur les réponses à toutes les questions d'évaluation.

8.3.3 Allouer les ressources

En concevant sa méthode, l'équipe d'évaluation tente de répartir ses ressources limitées de façon adéquate entre les questions. Certaines questions méritent d'être traitées avec des outils coûteux comme une enquête par questionnaire auprès des utilisateurs finaux, plusieurs études de cas, un *focus group*, etc. D'autres questions peuvent être traitées sur la base d'une simple analyse documentaire et de quelques entretiens au sein de la Commission européenne et du Gouvernement du pays partenaire.

Il est également raisonnable d'investir des ressources substantielles dans une question qui soulève des problèmes de faisabilité.

Une question est rarement inévaluable dans l'absolu. Les problèmes de faisabilité proviennent plutôt d'une accumulation de

difficultés et de contraintes. Au début de l'évaluation, il est souvent possible d'amender une question difficile afin de la rendre plus évaluable. Cela peut être fait, par exemple, en limitant le champ de la question ou en choisissant de l'appliquer à un effet moins éloigné ou à un effet probable si l'effet réel n'est pas encore observable. Une fois la question validée, l'équipe d'évaluation doit concevoir une méthode appropriée et allouer des ressources adéquates.

8.3.4 Coût et contraintes de temps

Les versions successives de la méthode sont conçues au sein de l'équipe d'évaluation jusqu'à ce que les contraintes suivantes soient respectées :

- La mise en œuvre des outils d'évaluation s'ajuste au calendrier du processus d'évaluation
- Le coût des outils d'évaluation (ressources humaines, coûts techniques, frais de déplacement et de séjour) s'ajuste au budget global de l'évaluation
- Une force de travail qualifiée est disponible sur le terrain de façon suffisante pour mettre en œuvre les outils d'évaluation de façon professionnelle.

8.4 Développement des outils

Le volume 4 présente une douzaine d'outils fréquemment employés, avec des exemples d'utilisation dans le contexte de l'évaluation de l'aide au développement.

Chaque outil implique une phase préparatoire comme l'écriture d'un guide d'entretiens, la rédaction et le test d'un questionnaire, etc. Le développement des outils se fait selon les étapes suivantes :

- Liste des questions et étapes de raisonnement que l'outil doit aider à traiter
- Spécifications techniques pour la mise en œuvre de l'outil
- Risques prévisibles qui peuvent compromettre ou affaiblir la mise en œuvre de l'outil et la manière de les gérer
- Modalités de compte-rendu au sein de l'équipe d'évaluation
- Modalités de compte-rendu dans le rapport final
- Responsabilités au sein de l'équipe d'évaluation dans la mise en œuvre de l'outil

- Critères de qualité et processus de contrôle qualité
- Calendrier d'exécution
- Ressources allouées.

9 Collecte d'information

RÉSUMÉ

Le plan de collecte d'information

En vue de répondre aux questions, l'équipe d'évaluation rassemble des informations qui sont déjà disponibles (données secondaires) et met en œuvre des outils de collecte en vue d'obtenir des informations nouvelles (données primaires).

Difficultés les plus fréquentes

Même si le plan de collecte d'information de terrain a été soigneusement préparé, l'équipe d'évaluation peut rencontrer des difficultés au cours de son travail de terrain, telles que : l'accès aux informateurs, le manque de coopération de la part des informateurs, la distance culturelle, l'absence ou la faiblesse d'une information recherchée.

Fiabilité des données collectées

Les travaux d'évaluation sont sujets à plusieurs sortes de risques. La fiabilité des données collectées peut être mise en danger par des biais tels que : biais de confirmation, biais d'empathie, autocensure, etc.

9.1 Le plan de collecte d'information

Le plan de collecte d'information finalise le travail de conception méthodologique (Cf. Section 8.5).

Les données d'enquête (primaires) sont collectées à l'aide d'outils tels que : entretiens, questionnaire, *focus group*, visites, observation directe, etc.

Pour éviter les duplications, les coûts inutiles et la lassitude des informateurs, il convient de s'appuyer le plus possible sur les informations existantes (données secondaires) telles que : données de gestion et de suivi, études et recherches relatives au domaine considéré, sources statistiques publiées, rapports d'évaluation et leurs annexes, etc.

Ces informations peuvent être obtenues à moindre coût. Elles peuvent contribuer à fournir des éléments de réponses à une partie des questions posées.

Si l'équipe d'évaluation oublie de mentionner des données existantes et pertinentes, son travail pourra facilement être critiqué par les parties prenantes qui considèrent ces informations comme importantes, et la crédibilité de l'évaluation sera amoindrie.

Canaux pour identifier et collecter des données secondaires

- Les questionnaires, les agences de mise en œuvre, les opérateurs et les partenaires
- Les experts du domaine considéré
- Internet
- Les offices statistiques et instances de suivi
- Les scientifiques et la littérature professionnelle.

9.2 Problèmes fréquents et solutions

Même si le plan de collecte d'information a été préparé judicieusement, l'équipe d'évaluation rencontre souvent des difficultés en réalisant son travail de terrain. Les sections ci-dessous exposent les problèmes les plus fréquemment rencontrés et la façon de les résoudre.

9.2.1 Accès aux informateurs

Le processus d'échantillonnage se révèle difficile :

- Décider si oui ou non un échantillon de taille réduite peut vraisemblablement fournir des résultats statistiquement valides.

Un informateur ne s'exprime pas librement :

- Centrer les entretiens sur des faits préférablement à des opinions
- Proposer de conserver l'information de manière anonyme et expliquer comment la confidentialité sera assurée.

Un informateur s'exprime d'une manière qui semble délibérément biaisée :

- Centrer les demandes d'information sur des faits, et non sur des opinions.

9.2.2 Différence culturelle

Un informateur ou une source d'information ne peut être consulté(e) que dans la langue locale :

- L'équipe d'évaluation doit comprendre au moins un membre qui parle couramment la langue locale (la traduction et l'interprétation génèrent toujours une perte d'information importante).

Il y a une grande distance culturelle entre l'équipe d'évaluation et la population enquêtée :

- L'équipe d'évaluation doit compter un ou plusieurs membres capables de faire le lien entre les deux cultures.

9.2.3 Manque ou faiblesse des données

Une source d'information se révèle incomplète :

- Si possible, extrapoler les données manquantes et croiser avec d'autres sources de données.

Une source d'information se révèle peu fiable :

- Si possible, comprendre les biais, ajuster les données et croiser avec d'autres sources.

Recommandations

Toute évaluation crée un sentiment d'incertitude, qui peut rendre certaines parties prenantes réticentes à coopérer, voire franchement hostiles. Dans un tel cas, prendre une attitude positive, insister sur les utilisations attendues de l'évaluation, promettre l'impartialité, et se concentrer sur les faits plutôt que sur les opinions.

Si une source d'information n'est pas accessible ou si une technique d'enquête n'est pas applicable, changer alors le plan de travail dans le but de collecter des informations similaires grâce à d'autres sources d'information.

Faire attention aux biais et risques affectant la fiabilité. S'efforcer de les comprendre et de les signaler.

Éviter de compter sur une unique source d'information pour faciliter le croisement de données au moment de l'analyse.

9.3 Fiabilité des données

En collectant l'information, l'équipe d'évaluation rencontre différents risques de biais qui peuvent affecter la fiabilité des données collectées.

Pourquoi faut-il considérer avec attention les biais ?

- Pour améliorer la fiabilité de la collecte de données
- Pour apprécier la qualité de l'évaluation
- Pour comprendre les limites des conclusions qui reposent sur des données de fiabilité faible

Biais de confirmation

Ce biais concerne l'ensemble des approches de collecte de données. Il résulte de la tendance à rechercher les preuves qui soient cohérentes avec la logique d'intervention, à la place de rechercher les preuves qui pourraient la réfuter.

L'équipe d'évaluation et les informateurs sont sujets à ce biais lorsqu'ils ont tendance à se focaliser sur les effets attendus et à

systématiquement oublier les facteurs externes, les effets non attendus, les effets négatifs, les interactions avec d'autres politiques, les parties prenantes externes, les options alternatives de mise en œuvre, etc.

Autocensure

Dans certains cas, les informateurs peuvent être réticents à répondre librement aux questions, simplement parce qu'ils ressentent un risque. Ils ont souvent tendance à exprimer strictement le point de vue de leur institution ou de leur hiérarchie.

Ce biais est traditionnellement combattu par une garantie de traitement confidentiel des réponses. La personne posant les questions doit insister sur des questions factuelles et éviter de collecter des opinions.

Stratégie des informateurs

Celui qui aura été impliqué dans l'intervention peut distordre une partie de l'information qu'il fournit afin d'attirer les conclusions de l'évaluation vers ses points de vue.

Ce biais pourra être réduit si le plan de collecte d'information couvre l'ensemble des parties prenantes et si différentes sources d'information sont croisées.

Question induisant les réponses

Ce biais et les suivants sont fréquents dans les entretiens et les questionnaires.

La manière dont sont posées les questions et également l'ordre des questions peuvent générer un biais positif ou négatif.

Ce biais sera limité si les questionnaires sont élaborés et testés par des professionnels expérimentés.

Biais d'empathie

Lors de ses entretiens, l'équipe d'évaluation cherche à créer une atmosphère conviviale (empathique) afin de faciliter la collecte d'information, d'obtenir un bon taux de réponse et/ou une finalisation rapide de l'enquête.

Ce phénomène introduit un biais positif systématique dans les réponses et a tendance à faire surestimer les bénéfices de l'intervention et sous-estimer le rôle des facteurs externes.

En outre, les personnes interrogées peuvent ne pas avoir une opinion préétablie à propos de la question qui leur est posée. Elles s'efforcent de se faire leur propre opinion en quelques secondes en répondant à l'enquêteur ou au questionnaire. Lorsqu'elles font cet exercice dans une atmosphère artificiellement positive, cela renforce le biais d'empathie.

Échantillon non représentatif

Ce biais peut être un sujet d'inquiétude si l'équipe d'évaluation génère des données quantitatives à l'aide d'une enquête par questionnaire. Il doit être également considéré lorsqu'on utilise des données secondaires issues d'une enquête par questionnaire. Dans ces cas, on doit vérifier que l'échantillon d'informateurs enquêtés est suffisamment large et représentatif de la population dans son ensemble.

Biais de sélection de l'échantillon

Les personnes qui acceptent d'être interrogées peuvent ne pas être représentatives de l'ensemble du public cible.

Ce biais peut être contrôlé en menant une enquête qualitative spéciale sur un petit nombre de non répondants, en sachant que cet exercice implique des coûts additionnels.

Recommandations

Mélanger systématiquement des questions positives et négatives afin de réduire le biais d'empathie et les biais sur les questions.

Être très crédible au moment de promettre la confidentialité pour limiter l'autocensure des répondants, et respecter strictement cette promesse.

Ne jamais compter sur une seule catégorie de parties prenantes (gestionnaires de programmes, bénéficiaires) afin de réduire les biais stratégiques.

10 Analyse

RÉSUMÉ

Stratégies d'analyse

Quatre stratégies d'analyse sont envisageables : analyse du changement, méta-analyse, analyse d'attribution et analyse de contribution. Les trois dernières permettent de répondre aux questions concernant les causes et les effets.

Déroulement de l'analyse

Lorsque la stratégie a été choisie et que les données ont été collectées, l'analyse suit tout ou partie des quatre stades suivants : traitement des données, exploration, explication, confirmation.

Validité de l'analyse

La validité est obtenue quand les constats ont une portée générale au-delà de l'échantillon observé ou des cas étudiés (validité externe), quand les constats découlent des informations collectées sans aucun biais (validité interne), et quand les informations collectées décrivent exactement ce qui était à décrire (validité de construction).

10.1 Stratégies d'analyse

10.1.1 Analyse du changement

Cette stratégie vise à préciser l'évolution des indicateurs mesurés et/ou qualifiés sur une période de temps. Elle ne cherche pas à établir une relation de causalité entre l'intervention évaluée et le changement.

Cas des questions causales

Les questions de causalité concernent les effets de l'intervention évaluée.

Ces questions font appel à une observation du changement puis à une attribution du changement observé à l'intervention ou à une analyse de la contribution de l'intervention au changement observé.

Les questions d'efficacité et d'impact tendent à être des questions causales dans la mesure où elles relient l'intervention évaluée (la cause) et ses effets. Les questions d'efficience et de durabilité sont indirectement des questions causales puisque des effets réels doivent d'abord être analysés avant d'être qualifiés d'efficaces ou de durables.

D'une manière générale, les questions de pertinence et de cohérence ne sont pas des questions causales.

10.1.2 Méta-analyse

Cette stratégie extrapole les constats d'autres évaluations et études, après avoir soigneusement vérifié leur validité et leur transférabilité.

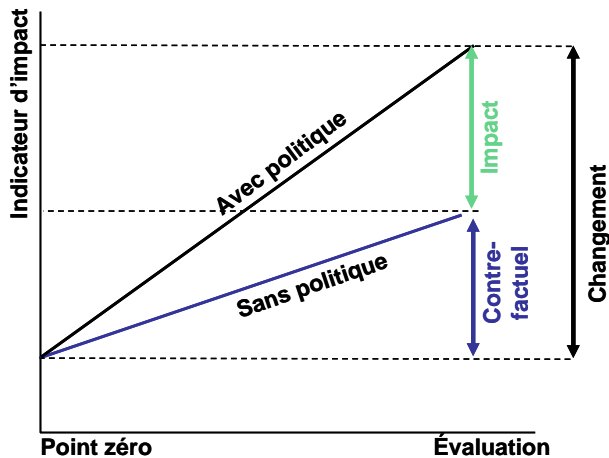
10.1.3 L'analyse d'attribution

L'analyse d'attribution a pour objet d'apprécier la proportion des changements observés qui sont effectivement attribuables à l'intervention évaluée. Elle implique la construction d'un scénario contrefactuel (Cf. Encadré 11).

La ligne « avec politique » montre les changements observés, mesurés avec un indicateur d'impact, entre le début de la période évaluée (point zéro) et la date de l'évaluation. Par exemple : l'emploi local a augmenté, l'illettrisme a reculé.

La ligne « sans politique », également appelée contrefactuelle, est une estimation de ce qui se serait passé sans l'intervention. Elle peut être obtenue par des approches appropriées comme les groupes de contrôle avec tirage au sort, les groupes de comparaison ou les techniques de modélisation. L'impact est apprécié à partir de l'indicateur « avec politique » tel qu'il a été observé, en soustrayant l'estimation de l'évolution « sans politique ».

Encadré 11 – Scénario contrefactuel



Puisqu'elle est dérivée d'un contrefactuel estimé, l'appréciation d'impact est elle-même une estimation. En d'autres termes, les impacts ne peuvent pas être mesurés directement. Ils peuvent juste être dérivés d'une analyse des indicateurs d'impact.

Seul un contrefactuel peut fournir une estimation quantitative de l'impact. Quand elle réussit, cette approche dispose d'un fort potentiel d'apprentissage et de rétroaction. Elle est cependant très exigeante en termes de données et de ressources humaines.

10.1.4 L'analyse de contribution

L'analyse de contribution vise à démontrer que l'intervention évaluée est une des causes des changements observés ou qu'elle ne l'est pas. Elle repose sur des chaînes d'arguments logiques qui sont vérifiés grâce à une analyse confirmatoire attentive.

Elle comprend les étapes successives suivantes :

- Raffiner les liens de cause à effet qui mettent en relation la conception et la mise en œuvre d'une part, et l'effet évalué d'autre part. Cette étape se fonde sur les explications disponibles concernant le domaine évalué. Ces explications dérivent du diagramme des effets escomptés réalisé lors de la première phase de l'évaluation, ou de l'expertise de

l'équipe d'évaluation, ou encore des analyses exploratoires.

- Rassembler des éléments de preuves relatifs à chaque lien de cause à effet, tels que les constats d'études similaires, les relations causales énoncées par les interviewés, et les éléments de preuves issus d'enquêtes approfondies.
- Rassembler des éléments de preuves relatifs aux explications alternatives (autres interventions, facteurs externes).
- Développer une chaîne d'arguments établissant que l'intervention a (ou n'a pas) apporté une contribution.
- Soumettre le raisonnement à une critique systématique jusqu'à ce qu'il soit suffisamment robuste.

10.1.5 Les facteurs externes

Les facteurs externes, également appelés contextuels, exogènes ou facteurs concurrents, entravent ou amplifient les changements attendus tout en étant indépendants de l'intervention elle-même.

Dans une évaluation donnée, les facteurs externes sont potentiellement nombreux et il est crucial de mettre en lumière les plus importants.

Recommandations

Ne pas essayer d'identifier tous les facteurs externes possibles lors de la phase de structuration de l'évaluation et de la clarification de la logique d'intervention. Ils sont simplement trop nombreux. Cette tâche ne devrait être entreprise que dans le cas d'une question évaluative donnée.

10.2 Le déroulement de l'analyse

Une analyse est nécessaire pour convertir les données en constats, qui font eux-mêmes l'objet d'un jugement pour être convertis en conclusions. L'analyse est principalement effectuée question par question, mais certaines analyses peuvent englober plusieurs questions.

10.2.1 Le traitement des données

Les données sont traitées au moyen d'opérations telles que le croisement, la comparaison, le regroupement, le listage, etc.

- Le croisement est l'utilisation de plusieurs sources ou types de données en vue d'établir un fait.
- La comparaison procède par construction de tableaux, de graphiques, de cartes et/ou de classements. Les données peuvent être comparées selon une ou plusieurs dimensions comme le temps, les territoires, les catégories administratives, les catégories socio-économiques.
- Le regroupement fonctionne en agrégeant les données selon des typologies prédéfinies.
- Le listage fonctionne en identifiant les différentes dimensions d'un besoin, d'un effet, etc.

Les constats provisoires émergent à ce stade de l'analyse. Les stades suivants ont pour but d'approfondir et de renforcer les constats.

10.2.2 Exploration

L'analyse exploratoire creuse systématiquement les données collectées afin de découvrir de nouvelles explications plausibles :

- Nouvelles catégories / typologies
- Facteurs explicatifs non prévus
- Facteurs favorisant / contraignant la durabilité
- Effets inattendus
- Nouvelles hypothèses de causalité.

L'analyse examine l'ensemble des données (quantitatives et qualitatives) afin d'identifier des structures, des différences, des contrastes, des similitudes et/ou des corrélations. Par exemple, l'équipe d'évaluation effectue :

- L'analyse transversale de plusieurs études de cas
- Des comparaisons statistiques recoupant des bases de données de gestion, des bases de données statistiques, et/ou les résultats d'une enquête par questionnaires
- Des comparaisons entre des entretiens, des questionnaires et des documents.

L'approche est systématique et ouverte. Les techniques de créativité collective sont indiquées.

10.2.3 L'explication

À ce stade, on s'assure d'avoir atteint une compréhension suffisante en termes de :

- Concepts, catégories et typologies définis avec précision
- Explications de causalité plausibles
- Identification des facteurs externes clés et des explications alternatives.

10.2.4 Confirmation

Le rôle de l'analyse confirmatoire est de vérifier les constats provisoires à travers une chaîne d'arguments valide et crédible.

Cette analyse vise à :

- Assurer que les constats sont de bonne qualité et qu'ils peuvent résister à n'importe quelle critique quand le rapport est édité
- Assurer la crédibilité des constats du point de vue des utilisateurs finaux
- Distinguer, dans le cas particulier des questions de causalité, les effets réels du changement observé.

Pour qu'un constat soit confirmé, il y a lieu de le soumettre à la critique par tous les moyens possibles, par exemple :

- Si le constat provient d'une analyse statistique, les tests de validité sont-ils concluants ?
- Si le constat a été suggéré par une étude de cas, est-il contredit par une autre ?
- Si le constat provient d'une enquête, peut-il être expliqué par un biais dans cette enquête ?

La force de la preuve

Les constats doivent résister à la critique. C'est pourquoi ils doivent reposer sur des éléments de preuve convergents tels que des faits, des rapports ou des témoignages. Quatre niveaux de preuves sont à considérer :

- **Fait observé**

Les preuves factuelles sont les plus fortes. Les faits observés peuvent prendre la forme de rapports de visite, de photographies, de rapports de gestion, ou n'importe quelle trace matérielle.

- **Témoignage vécu**

Il s'agit également d'une preuve forte, par exemple les bénéficiaires directs déclarent avoir changé leur attitude après avoir participé à un programme.

- **Fait approché (*proxy*)**

Une preuve factuelle indirecte est aussi appelée preuve circonstancielle ou indicateur approché (*proxy*), par exemple, durant les derniers mois, plusieurs concurrents d'une entreprise aidée ont fait faillite, ce qui indique que le niveau d'aide était excessif et a distordu la concurrence.

La force de ce type de preuve dépend avant tout de la force du raisonnement logique qui fonde la déduction.

- **Témoignage rapporté**

Les témoignages indirects sont les types de preuves les plus faibles, par exemple des gestionnaires de programmes déclarent que les entreprises bénéficiaires ont fortement amélioré leur compétitivité.

La force de ce type de preuves dépend avant tout de l'autorité de l'expert qui témoigne.

- Si le constat est basé sur une source d'information, est-il contredit par une autre source ?
- Le constat est-il relatif à un changement qui peut être expliqué par des facteurs externes qui auraient été oubliés par l'équipe d'évaluation ?

- Le constat contredit-il les opinions des experts ou les leçons apprises ailleurs ? Si oui, cela peut-il être expliqué ?
- Les membres du groupe de référence de l'évaluation ont-ils des arguments qui contredisent le constat ? Si oui, ces arguments sont-ils justifiés ?

Suggestions pour l'analyse

Consacrer des interactions relativement longues à la discussion du rapport final afin de permettre une analyse confirmatoire minutieuse. S'assurer que l'équipe d'évaluation a réservé des ressources suffisantes dans ce but.

Tous les constats n'exigent pas le même niveau de confirmation. Concentrer les efforts sur les constats qui soutiennent les conclusions les plus contestées, les leçons les plus susceptibles d'être transférées, ou les recommandations les plus difficiles à accepter.

Afin d'augmenter la crédibilité de l'évaluation, il est intéressant de présenter en annexe les critiques auxquelles les constats ont résisté pendant l'analyse confirmatoire.

10.3 Validité de l'analyse

La validité est une qualité essentielle de l'analyse. Elle est obtenue quand :

- Les constats ont une portée générale au-delà de l'échantillon observé ou des cas étudiés (validité externe)
- Les constats découlent des informations collectées sans aucun biais (validité interne)
- Les informations collectées décrivent exactement ce qui était à décrire (validité de construction).

10.3.1 Validité externe

La validité externe est obtenue lorsqu'une méthode d'évaluation aboutit à des constats généralisables à d'autres groupes, d'autres zones, d'autres périodes, etc. Elle est atteinte lorsque l'équipe d'évaluation peut montrer qu'une intervention similaire dans un autre contexte aurait eu les mêmes effets, selon des conditions données.

10.3.2 Validité interne

La validité interne correspond à la qualité d'une méthode d'évaluation qui limite autant que possible les biais imputables à la collecte et au traitement des données. La validité interne est obtenue lorsque l'équipe d'évaluation fournit des arguments indiscutables montrant que les constats dérivent de faits.

10.3.3 Validité de construction

La validité de construction est obtenue lorsque les concepts clés sont clairement définis et que les indicateurs reflètent ce qu'ils sont censés refléter.

11 Jugement

RÉSUMÉ

Conclusions et leçons

Les conclusions apportent des réponses claires aux questions posées au début de l'évaluation. Elles intègrent les jugements sur la valeur et les mérites de l'intervention.

Les leçons sont des conclusions transférables aux prochains cycles de la même intervention, à d'autres interventions ou à des interventions similaires dans d'autres contextes.

Recommandations

Les recommandations ne comportent pas de jugement à proprement parler, mais elles dérivent des conclusions. Elles visent à améliorer ou réformer l'intervention évaluée, ou à préparer la conception d'une nouvelle intervention pour le cycle suivant.

11.1 Conclusions et leçons

Les conclusions incluent un jugement

Autant que possible, le rapport d'évaluation énonce séparément les constats [qui découlent uniquement des faits et des analyses], et les conclusions [qui comportent un jugement sur les mérites de l'intervention].

Pour formuler ses conclusions, l'équipe d'évaluation fait référence aux critères de jugement (également appelés critères d'appréciation raisonnée) convenus lors de la phase de documentation de l'évaluation (*desk*) (Cf. Section 7.1). Lorsque cela est possible, les conclusions explicitent les niveaux cibles (*targets*).

Cette démarche renforce la qualité dans la mesure où elle oblige l'équipe d'évaluation à expliciter la façon dont elle formule son jugement. Le fait de reconnaître que les conclusions comportent un jugement de valeur conduit aussi à mieux respecter les principes éthiques présentés dans l'encadré ci-après.

Éthique du jugement

• Responsabilité

Dans la phase de synthèse, l'équipe d'évaluation applique les critères de jugement convenus de façon aussi fidèle que possible et produit ses propres conclusions. Les conclusions sont discutées au sein du groupe de référence mais restent de la seule responsabilité de l'équipe d'évaluation.

• Légitimité

Les questions et critères prennent en compte les besoins et le point de vue de l'institution publique qui a initié l'évaluation.

Les membres du groupe de référence peuvent apporter des points de vue différents, ce qui renforce la légitimité de l'évaluation.

L'équipe d'évaluation conduit des entretiens qui lui permettent d'identifier éventuellement d'autres points de vue qui ne s'expriment pas au sein du groupe de référence et qui sont pourtant légitimes. Elle en fait part aux réunions du groupe de référence et elle en tient compte.

• Impartialité

Les acteurs font en sorte que toutes les opinions soient entendues, même celles qui ne sont pas portées par les voix les plus puissantes ou les plus nombreuses. Ils sont conscients des relations de pouvoir asymétriques et corrigent les biais qui en résultent.

Les membres de l'équipe d'évaluation sont familiers et respectueux des croyances, mœurs et coutumes des groupes concernés.

Il est rendu compte des divergences de vues dans le rapport et dans les réunions du groupe de référence.

• Protection des personnes

Les conclusions portent sur les mérites de l'intervention évaluée, pas sur les personnes qui la mettent en œuvre ou qui en bénéficient.

Au stade du rapport final provisoire, l'équipe d'évaluation peut être amenée à affiner les critères de jugement et les niveaux cibles.

Dans ce cas, il convient de le préciser et d'en discuter avec le groupe de référence.

Réponses aux questions et conclusions générales

Les conclusions apportent des réponses claires aux questions posées au début de l'évaluation.

Un chapitre du rapport d'évaluation synthétise ces réponses sous la forme de conclusions d'un niveau plus général, de façon à porter une appréciation d'ensemble sur l'intervention évaluée. Cette synthèse ne suit pas l'ordre des questions ni celui des critères d'évaluation (efficacité, efficience, cohérence, etc.).

Récapitulation des conclusions

Le rapport contient une liste des conclusions faisant référence à la façon dont elles dérivent des faits, des analyses et des critères de jugement. Le rapport contient également une auto-appréciation des limites méthodologiques qui peuvent restreindre la portée ou l'utilisation de certaines conclusions.

Le rapport identifie les leçons, qui sont des conclusions transférables aux prochains cycles de la même intervention, à d'autres interventions ou à des interventions similaires dans d'autres contextes.

Un paragraphe ou un sous-chapitre reprend, en évitant toute répétition, les trois ou quatre conclusions majeures organisées par ordre d'importance. Cet exercice permet de mieux faire ressortir les messages de l'évaluation qui s'adressent aux responsables politiques au sein de la Commission.

Attention !

L'équipe d'évaluation et les personnes en charge de l'assurance qualité ont intérêt à relire attentivement le rapport final pour éliminer tous les jugements de valeur non indispensables ou involontaires.

L'équipe d'évaluation expose ses conclusions de façon balancée, sans privilégier de façon systématique les conclusions positives ou négatives.

11.2 Recommandations

11.2.1 Comment les élaborer et les présenter ?

Les recommandations doivent être liées aux conclusions sans être leur calque. Une recommandation dérive directement d'une ou plusieurs conclusions.

Attention !

Si une recommandation ne dérive pas clairement des conclusions, elle reflète probablement des idées préconçues ou les intérêts tactiques d'une des parties prenantes. Sa présence dans le rapport risque de discréditer l'évaluation.

Les recommandations sont regroupées et classées par ordre de priorité. Le rapport mentionne à qui elles sont destinées, par exemple : la Délégation de la Commission européenne, les services en charge de la conception de la prochaine intervention, etc.

Les recommandations sont utiles, opérationnelles et faisables. Les conditions de leur mise en œuvre sont précisées. Chaque fois que cela est possible et pertinent, les principales recommandations sont présentées sous formes d'options avec les conditions liées à chaque option ainsi que les conséquences prévisibles de la mise en œuvre de chaque option. Les recommandations sont présentées dans un chapitre spécifique qui met en valeur les recommandations dérivées des trois ou quatre principales conclusions.

11.2.2 Comment les promouvoir ?

Les recommandations n'ont de valeur que si elles sont entendues, et si possible intégrées par leurs destinataires. Pour promouvoir cette appropriation, le gestionnaire prépare une fiche contradictoire qui permet de :

- Lister les recommandations sous forme résumée
- Recueillir la position des services destinataires par rapport à ses recommandations
- Informer sur les suites effectivement données aux recommandations.

12 Assurance qualité

RÉSUMÉ

L'assurance qualité consiste à établir des points de contrôle de la qualité aux principales phases du processus, à préciser qui est responsable de la qualité et à établir les règles du jeu en cas de défaut de qualité.

L'assurance qualité renforce les conclusions de l'évaluation aux yeux des utilisateurs en montrant qu'il s'agit d'un travail impartial et rigoureux, respectant les standards professionnels.

La qualité se construit progressivement de façon à éviter de constater un défaut de qualité au stade du rapport final.

Les règles d'assurance qualité clarifient les relations entre le responsable de l'évaluation, le groupe de référence et l'équipe d'évaluation.

12.1 Les règles du jeu

Les règles de l'assurance qualité sont précisées dans les termes de références et rappelées lors de l'engagement de l'équipe d'évaluation.

Les éléments de qualité successifs sont sécurisés à chaque étape de façon à avancer sur des bases solides vers l'étape suivante.

Les règles concernent :

- L'approbation des documents
- Les critères de qualité
- La diffusion de l'appréciation de qualité du rapport final.

Le processus d'assurance qualité bénéficie de la contribution de tous les acteurs tout en limitant les conflits potentiels qui pourraient survenir entre eux, par exemple :

- Les acteurs qui détiennent les responsabilités dans l'intervention évaluée (les évalués) sont en bonne position pour apprécier la qualité des données collectées. Leur appréciation peut cependant être sensible au biais de confirmation. Ils doivent donc avoir un rôle consultatif, par

exemple au moyen de leur participation au groupe de référence.

- Le *leader* de l'équipe d'évaluation a un rôle majeur pour assurer la qualité, particulièrement en concevant une méthode adéquate, et en vérifiant la précision et l'impartialité du rapport. Il risque toutefois de survaloriser les points de vue des parties prenantes influentes, et spécialement celles susceptibles de lui commander d'autres évaluations à l'avenir. L'appréciation qualité du gestionnaire de l'évaluation contribue à prévenir ce risque.

Les critères de qualité

La qualité du rapport d'évaluation est notée selon neuf critères à l'aide d'une grille qui comprend cinq niveaux : Excellent, Très bon, Bon, Médiocre et Inacceptable.

Critère 1 : Satisfaction des demandes

Le rapport traite-t-il de façon adéquate les demandes d'information formulées par les commanditaires et correspond-il aux termes de référence ?

Critère 2 : Pertinence du champ

Le rapport fait-il un examen complet de la raison d'être de l'intervention examinée, de ses réalisations, résultats et impacts, ainsi que de ses interactions prévues et imprévues avec d'autres politiques et leurs conséquences ?

Critère 3 : Justification de la méthode

La conception de l'évaluation est-elle adaptée et adéquate pour fournir l'ensemble des constats qui serviront à répondre aux principales questions de l'évaluation, compte tenu des limites méthodologiques ?

Critère 4 : Fiabilité des données

Les données primaires et secondaires sélectionnées sont-elles adaptées ? Sont-elles suffisamment fiables pour l'usage qui en est attendu ?

Critère 5 : Solidité de l'analyse

Les informations quantitatives et qualitatives sont-elles analysées de façon adéquate et systématique selon les règles de l'art, de sorte qu'elles permettent de répondre aux questions de l'évaluation de façon valide ?

Critère 6 : Crédibilité des constats

Les constats découlent-ils de façon logique et justifiée d'une analyse des données et d'interprétations basées sur des hypothèses et sur une logique présentées avec soin ?

Critère 7 : Validité des conclusions

Le rapport fournit-il des conclusions claires ? Sont-elles basées sur des constats crédibles ?

Critère 8 : Utilité des recommandations

Les recommandations sont-elles équitables, non biaisées en faveur de points de vue personnels ou de parties prenantes, et assez détaillées pour être concrètement applicables ?

Critère 9 : Clarté du rapport

Le rapport décrit-il clairement l'intervention évaluée, y compris son contexte et sa finalité, ainsi que le processus et les constats de l'évaluation de façon telle que les informations fournies soient aisément compréhensibles ?

Une grille simplifiée est présentée dans le Volume 3 pour l'évaluation des projets et programmes.

Les critères 1 et 2 sont vérifiés au stade du rapport de démarrage.

Au stade du rapport de la phase de documentation, il convient de vérifier les critères 1 à 3, et le critère 4 pour ce qui est des données déjà collectées.

L'ensemble des critères 1 à 9 est apprécié au stade du rapport final (y compris le projet de rapport final).

Sur le site Web : conseils pour apprécier la qualité selon chacun des neuf critères

12.2 L’approbation des rapports

L’approbation des rapports consiste en la reconnaissance formelle par le gestionnaire de l’évaluation qu’un document remis a un contenu adéquat et satisfait aux critères de qualité applicables en terme de méthodologie.

12.2.1 Quelles sont les étapes de l’approbation ?

Sauf exception, l’approbation d’un document se déroule de la façon suivante :

- Le responsable vérifie que le document a bien la forme et le contenu requis, et qu’il ne comporte pas de défaut de qualité majeur. Dans le cas contraire, il demande immédiatement une nouvelle version.
- Le responsable recueille l’avis des personnes ou services qui doivent être consultés, soit au cours d’une réunion (groupe de référence), soit par courriel en indiquant un délai limité au-delà duquel la prise en compte des commentaires n’est pas garantie.
- Le responsable liste ses propres demandes d’amélioration de qualité, attache les autres demandes reçues, distingue les demandes d’ordre méthodologique et les demandes de fond, et transmet l’ensemble des remarques à l’équipe d’évaluation.
- L’équipe d’évaluation prend en compte toutes les demandes d’amélioration, soit en modifiant son document en conséquence, soit en mentionnant en annexe comment la demande sera traitée à un stade ultérieur, soit en expliquant en annexe pourquoi la demande n’a pas été traitée.
- Une nouvelle version du document est rapidement présentée au responsable qui l’accepte ou la refuse.
- Lorsqu’elle est requise, l’approbation conditionne le passage à l’étape suivante.

12.2.2 Quels documents sont approuvés ? Par qui ?

Les documents soumis à approbation sont les suivants :

- Rapport de démarrage (*inception*)
- Rapport de la phase de documentation (*desk*)
- Rapport final provisoire

- Rapport final.

L'approbation est généralement donnée par le gestionnaire de l'évaluation et éventuellement confirmée par son supérieur hiérarchique.

Glossaire

Le vocabulaire et la terminologie développés dans les quatre guides correspondent aux définitions de l'OCDE d'une part et de la Commission européenne d'autre part.

Définitions de l'OCDE

En 2002, l'OCDE a publié un glossaire des principaux termes relatifs à l'évaluation et à la gestion axée sur les résultats (Comité d'aide au développement - Groupe de travail sur l'évaluation de l'aide). Rédigé en anglais, français et espagnol, il a ensuite été traduit dans d'autres langues. Il participe à l'effort commun de la communauté de l'aide au développement, y compris la Commission européenne, en vue de réduire la confusion terminologique et de promouvoir un vocabulaire commun.

Définitions de la Commission européenne

Le Règlement Financier impose un cadre applicable à toutes les politiques de la Commission. Ce cadre a été décliné dans des guides d'évaluation et un glossaire rédigé par la Direction générale du budget.

Comparaison des définitions

Les définitions de l'OCDE et celles de la Commission européenne ont une égale légitimité dans le contexte de l'évaluation de l'aide extérieure européenne. Une comparaison approfondie a été menée, avec pour résultat que : **presque tous les termes sont définis de façon cohérente.**

Seuls quatre couples de définitions sont suffisamment différents pour provoquer des risques de malentendu :

Risque de malentendus (quatre termes)		
	OCDE	CE
Impact	... effets à long terme produits par une intervention de développement ...	Un terme général utilisé pour décrire les effets d'une intervention dans la société ...
Outcome (traduit par effet direct)	Les effets, potentiels ou avérés à court ou à moyen terme, des réalisations d'une intervention	Les impacts à plus long terme, habituellement exprimés en terme de conséquences socio-économiques au sens large
Résultat	... Les réalisations, effets directs ou impacts d'une intervention de développement (attendus ou non, positifs et/ou négatifs).	L'impact initial d'une intervention...
Efficience	Mesure de la façon dont les ressources économiques (fonds, expertise, temps etc.) sont converties en résultats (selon la définition des résultats de l'OCDE, l'efficience peut concerner les réalisations mais aussi tout niveau d'effet)	Mesure dans laquelle les effets produits le sont à un coût raisonnable (une définition qui ne concerne pas les réalisations)

Les choix suivants ont été faits pour rédiger ce guide :

Options choisies		
	Glossaire de base	Définition
Impact	OCDE	... effets à long terme produits par une intervention de développement
Outcome (effet direct)	Le terme n'a été ni utilisé ni défini	
Résultat	CE	L'impact initial d'une intervention ...
Efficience	Mixte (basé sur l'OCDE)	La mesure dans laquelle les réalisations et/ou les effets désirés sont atteints en utilisant le moins de ressources possible (fonds, expertise, temps, coûts administratifs, etc.)

Dans d'autres documents publiés par la Commission européenne, le terme résultat peut être employé au sens large, comme le définit l'OCDE/CAD, par exemple : Suivi orienté vers les résultats.

Disponible sur le site Web Europa

http://ec.europa.eu/europeaid/index_fr.htm

De nombreux exemples et *check-lists*, notamment :

- Exemples de questions pour une évaluation au niveau pays, au niveau région, pour une évaluation thématique ou sectorielle
- Exemples d'articulation entre questions, critères de jugement et indicateurs
- Exemple de tableau de conception et de développement d'un outil
- Grille d'appréciation de la qualité et indications pour remplir la grille
- Exemples de résumés et articles présentant une évaluation.

Liste des encadrés

Encadré 1 – Interventions simples / complexes.....	17
Encadré 2 - Distinction entre secteurs et aspects transversaux...	19
Encadré 3 – Évaluation concernant plusieurs cycles	23
Encadré 4 – Calendrier et rôle du groupe de référence	33
Encadré 5 – Diagramme des effets attendus.....	41
Encadré 6 – Focaliser l'évaluation	44
Encadré 7 – Principaux critères d'évaluation	50
Encadré 8 – Indicateurs et logique d'intervention	62
Encadré 9 – De la question à l'indicateur : exemple	63
Encadré 10 – Structure d'un tableau de conception	67
Encadré 11 – Scénario contrefactuel.....	82

COMMISSION EUROPEENNE

Méthodologie d'évaluation de l'aide extérieure de la Commission européenne

2006 – 101 pp. – 21 x 14,8 cm

Luxembourg : Office des publications officielles des Communautés européennes

ISBN : 92-79-00682-7