

L'évaluation pour un vrai changement : **Le point de vue des praticiens humanitaires**

Alistair Hallam et Francesca Bonino

ALNAP est un réseau d'apprentissage qui soutient le secteur humanitaire afin d'améliorer la performance des acteurs via l'apprentissage, le partage entre pairs et la recherche. ALNAP est un réseau unique dont les membres sont des acteurs issus de l'ensemble du secteur humanitaire : bailleurs de fonds, ONG, Mouvement international de la Croix-Rouge et du Croissant-Rouge, agences des Nations unies, consultants indépendants et institutions académiques. www.alnap.org/using-evaluation

Remerciements

Franziska Orphal et Alexandra Warner du Secrétariat d'ALNAP ont offert un appui inestimable tout au long de ce projet en matière de communication et d'assistance à la recherche. Paul Knox-Clarke et Patricia Curmi ont respectivement aidé à réviser les premières versions de cette étude et contribué au travail d'élaboration du contenu.

Merci également à tous ceux qui ont apporté leur éclairage, des informations pour les études de cas et, plus précieux encore, leur temps pour débattre de tous ces sujets avec nous. La liste des contributeurs (située à la fin de ce rapport) comprend les personnes interrogées et toutes celles qui ont apporté leur regard d'évaluateur via la communauté de pratique de l'évaluation d'ALNAP. Toute erreur d'interprétation ou lacune de ce texte relève de la seule responsabilité des auteurs.

Enfin, nous souhaitons remercier tous ceux qui ont aidé à la conception de la première version de ce document, en particulier : John Mitchell, Ben Ramalingam, Yuka Hasegawa, Rachel Bedouin, Aine Hearn et Karen Proudlock ainsi que les nombreux membres – ou non – d'ALNAP qui ont accepté d'être interrogés durant la phase initiale de ce projet.

Citation suggérée

Hallam, A. and Bonino, F. (2013) *L'évaluation pour un vrai changement : Le point de vue des praticiens humanitaires*. ÉTUDE ALNAP. Londres: ALNAP/ODI.



Attribution — Vous devez citer ce travail de la manière indiquée par l'auteur ou le titulaire des droits (mais en aucun cas d'une manière pouvant suggérer qu'ils vous soutiennent ou appuient votre utilisation de l'œuvre).



Non commercial — Vous n'êtes pas autorisé à utiliser ce travail à des fins commerciales.



Share Alike — Si vous modifiez, transformez ou adaptez ce travail, vous n'avez le droit de distribuer la création qui en résulte que sous la même licence ou sous une licence identique à celle-ci.

Ce travail est sous licence Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Pour plus d'informations sur cette licence, veuillez consulter ce lien : creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US.

Conçu par Deborah Durojaiye
Revu par Nina Behrman
Traduit par Pierre-Julien Brunet (Groupe URD)

Cette publication est aussi disponible en français sur le site web d'ALNAP au www.alnap.org/using-evaluation. Si vous avez des commentaires ou des suggestions, veuillez nous envoyer un courriel au aha@alnap.org.

À propos des auteurs

Alistair Hallam est économiste et médecin, avec plus de 20 années d'expérience dans le secteur humanitaire. Co-fondateur et directeur de Valid International, il a beaucoup écrit sur les questions humanitaires et est notamment l'auteur du guide de référence en matière d'évaluation des programmes d'aide humanitaire : *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*.

Francesca Bonino est chargée de recherche et s'occupe des questions d'évaluation, d'apprentissage et de redevabilité. Elle est titulaire d'une thèse portant sur le développement de l'évaluation dans le secteur humanitaire et au sein des organisations humanitaires.

Sommaire

INTRODUCTION	6
Contexte	6
Approche de l'étude et remarque méthodologique	9
Objectifs de l'étude	12
Un cadre pour exploiter au mieux les capacités de l'évaluation des actions humanitaires	13
DOMAINE DE CAPACITÉS 1 : LEADERSHIP, CULTURE, STRUCTURE ET RESSOURCES	18
Garantir que le leadership soutient l'évaluation et le suivi	19
Promouvoir une culture de l'évaluation	23
Réduire le risque que l'évaluation soit perçue comme une critique et les évaluateurs comme des critiques	26
Utiliser du personnel à la fois interne et externe pour réaliser des évaluations	29
Renommer les évaluations	31
Être flexible et prendre du plaisir à ce que l'on fait	33
Trouver les bonnes incitations	35
Créer des structures organisationnelles qui promeuvent l'évaluation	38
Sécuriser les ressources adéquates (financières et humaines)	40
DOMAINE DE CAPACITÉS 2 : OBJECTIF, DEMANDE ET STRATEGIE	43
Clarifier l'objectif de l'évaluation (redevabilité, audit, apprentissage) et l'articuler avec les politiques d'évaluation	44
Augmenter la demande en matière d'informations sur l'évaluation	48
Viser l'implication des parties prenantes	50

Garantir que les processus d'évaluation se déroulent au bon moment et font partie intégrante du cycle de prise de décision	53
Développer une approche stratégique pour sélectionner ce qui devrait être évalué	57
DOMAINE DE CAPACITÉS 3 : LES PROCESSUS ET SYSTÈMES D'ÉVALUATION	61
Renforcer les processus préévaluation	62
Améliorer la qualité des évaluations	64
Limiter les objectifs de l'évaluation	64
Impliquer les bénéficiaires	65
L'assurance qualité	66
S'engager dans une revue entre pairs de la fonction Évaluation	68
Diffuser les résultats de façon efficace	70
Renforcer les processus de suivi et post-évaluation, y compris les relations entre l'évaluation et la gestion plus générale des connaissances	75
Garantir qu'il existe une réponse de la direction aux évaluations	78
Garantir un accès et une facilité de recherche vis-à-vis de l'information évaluative	80
Réaliser des méta-évaluations, des synthèses d'évaluation, des méta-analyses et des bilans de recommandations	82
Observations finales	86
Liste des contributeurs	90
Bibliographie	91

Encadrés

1.	Une communauté de pratique pour la consultation et l'apprentissage entre évaluateurs	12
2.	Les différentes utilisations de l'évaluation	14
3.	L'importance du <i>leadership</i> dans l'utilisation des évaluations	19
4.	Le processus d'Action contre la Faim pour collecter et valider les bonnes pratiques	28
5.	Quels facteurs influencent la demande des décideurs en faveur de données d'évaluation et de leur utilisation ?	49
6.	Qui compose le public de votre évaluation ?	51
7.	Une évaluation est-elle toujours justifiée? Quelques éclairages de Carol Weiss	57
8.	Améliorer le ciblage des évaluations au sein de l'UNICEF	59
9.	Tirer le maximum des ateliers d'évaluation inter-acteurs sur le terrain	65
10.	Les trois grands principes de l'évaluation	69
11.	Sept nouvelles manières de communiquer les résultats d'une évaluation	72
12.	Les expérimentations d'USAID en matière de films et les apprentissages réalisés à partir du travail de terrain des évaluations	73
13.	Choisir des stratégies de diffusion de l'évaluation	74
14.	Lier les processus d'évaluation et de suivi	77
15.	Des exemples d'amélioration de l'accès à l'information en matière d'évaluation	80
16.	Le projet d'apprentissage des réponses d'urgence d'ACF	84
17.	Améliorer l'utilisation des résultats d'évaluation au sein du PAM	85

Regards d'évaluateurs :



1.	Focaliser l'évaluation sur des questions provenant de la direction générale	20
2.	Un changement au sein de la direction, une opportunité pour l'évaluation ?	21
3.	Le soutien de la direction vis-à-vis de l'évaluation : un exemple issu de l'évaluation du Cluster Global Education	22
4.	Oxfam Intermón formule ses critiques de façon moins personnelle	26
5.	Le Groupe URD a adopté un « processus d'évaluation itérative »	26
6.	L'expérience de l'UNHCR avec des équipes d'évaluation mixtes	31
7.	La flexibilité n'est pas un crime	34
8.	La flexibilité en matière de conception d'évaluation	34
9.	L'exemple de Tearfund pour intégrer l'évaluation à l'ensemble des services de l'organisation	39

10.	L'initiative de DFID pour augmenter le soutien à l'évaluation	41
11.	Être conscient des lignes de faille entre apprentissage et redevabilité	44
12.	Qui est responsable en matière de gestion et de redevabilité ?	47
13.	L'intérêt du DEC pour l'apprentissage tiré des réponses antérieures	48
14.	Le cycle de planification des évaluations au sein de Sida	50
15.	L'expérience du DEC pour renforcer l'appropriation du travail d'évaluation par les parties prenantes	51
16.	Tirer le maximum de l'implication du bureau pays dans l'évaluation	52
17.	L'expérience du PAM en matière d'intégration de l'évaluation dans le développement de politiques	56
18.	L'assurance qualité et le système de supervision des évaluations de l'UNICEF	68
19.	Le grand impact des rapports de fin de mission	72
20.	Les ateliers post-mission du DEC	79

Acronymes et abréviations

ACDI	Agence canadienne de développement international	DEC	Disasters Emergency Committee
ACF	Action contre la faim	DECAF	DEC Accountability Framework
ALNAP	Active Learning Network for Performance and Accountability in Humanitarian Action (Réseau d'apprentissage actif pour la performance et la redevabilité dans l'action humanitaire)	DFID	UK Department for International Development (Département du développement international du gouvernement britannique)
BM	Banque mondiale	EAH	Évaluation de l'action humanitaire
CAD	Comité d'aide au développement (OCDE)	ECB	Emergency Capacity Building Project (Renforcement des capacités d'urgence)
CdP	Communauté de pratique	ECB-AIM	ECB Accountability and Impact Measurement (mesure de la redevabilité et de l'impact du projet ECB)
CR/CR	Mouvement Croix-Rouge et Croissant-Rouge	EQAS	Evaluation Quality Assurance System (Système d'assurance qualité des évaluations du PAM)
DARA	Development Assistance Research Associates		

ERD	Base de données des rapports d'évaluation (ALNAP)	ODI	Overseas Development Institute
ETR	Evaluation en temps réel	ONG	Organisation non gouvernementale
FAO	Organisation des Nations unies pour l'alimentation et l'agriculture	ONU	Organisation des Nations unies
FICR	Fédération internationale des Sociétés de la Croix-Rouge et du Croissant-Rouge	PAM	Programme alimentaire mondial des Nations unies
GEROS	Global Evaluation Reports Oversight System (Système global de supervision des rapports d'évaluation de l'UNICEF)	PNUD	Programme des Nations unies pour le développement
Groupe URD	Groupe Urgence, Réhabilitation, Développement	Sida	Département du ministère suédois des Affaires étrangères
HAP	Humanitarian Accountability Partnership	TdR	Termes de référence
IA-RTE	IASC Inter-Agency Real Time Evaluation (Évaluation en temps réel inter-agences du IASC)	UNEG	Groupe d'évaluation des Nations unies
IASC	Inter-Agency Standing Committee (Comité permanent inter-organisations)	UNHCR	Agence des Nations unies pour les réfugiés
ICAI	Commission indépendante pour l'impact de l'aide	UNICEF	Fonds des Nations unies pour l'enfance
IEG	Groupe indépendant d'évaluation (Banque mondiale)	UNRWA	Office de secours et de travaux des Nations unies pour les réfugiés de Palestine dans le Proche-Orient
S&E	Suivi et évaluation	USAID	Agence des États-Unis pour le développement international
MSF	Médecins sans frontières		
OCHA	Bureau de la coordination des affaires humanitaires des Nations unies		
OCDE	Organisation de coopération et de développement économiques		

INTRODUCTION

Contexte

L'évaluation de l'action humanitaire (EAH) est un domaine de pratiques au cœur du travail d'ALNAP. À ce sujet, ALNAP s'est concentré ces dernières années sur l'élaboration de guides (Cosgrave *et al.*, 2009 ; Beck, 2006 ; Proudlock et Ramalingam, 2009 ; Beck et Buchanan-Smith, 2008 ; Buchanan-Smith et Cosgrave, 2013) ainsi que sur des sujets en lien avec le suivi plus large de la performance au niveau systémique (voir par exemple : Beck, 2002 et 2003 ; Wiles, 2005 ; Beck et Buchanan-Smith 2008 ; Taylor, 2012). Les membres d'ALNAP ont toujours montré un réel intérêt pour dialoguer et s'engager dans des sujets d'actualité comme l'évaluation conjointe, l'évaluation d'impact ou encore les évaluations en temps réel en contexte humanitaire. De plus, l'étude référence de Peta Sandison sur l'utilisation de l'évaluation (Sandison, 2006) a ouvert un nouveau courant important de recherche et d'engagement conjoints avec les membres d'ALNAP sur le thème plus vaste de l'utilisation de l'évaluation (Hallam, 2011).

L'étude de Sandison s'est appuyée sur les ouvrages de référence en matière d'utilisation des évaluations (Chelimsky, 1997 ; Pawson et Tilley, 1997 ; Weiss, 1998 ; Guba et Lincoln, 1989 ; Feinstein, 2002 ; Patton, 2008). Pour la première fois, elle a attiré l'attention du secteur humanitaire au sens large sur les sujets, les opportunités et les défis liés au fait d'utiliser l'évaluation pour améliorer les pratiques humanitaires. S'appuyant sur ce précédent travail, la présente étude vise à faire en sorte que les membres d'ALNAP clarifient et rendent explicites certaines expérimentations pratiques et réflexions parmi les plus récentes dans le but d'augmenter les chances que l'évaluation soit utilisée de façon plus efficace pour contribuer à l'apprentissage et au changement des pratiques humanitaires.

L'importance accordée à l'amélioration de la conception, de l'approche et de la méthodologie des évaluations a été largement discutée dans la littérature (House, 1983 ; Scriven, 1991 ; Grasso, 2003 ; Kellogg Foundation, 2007 ; Rogers, 2009 ; Stern, 2012 ; WB-IEG, 2009). Cependant, les membres d'ALNAP ont eu le sentiment que la production des rapports d'évaluation de grande qualité n'était que la moitié du défi à relever pour

les acteurs humanitaires. L'autre moitié concerne le renforcement des capacités des individus, des équipes et des organisations à planifier, commanditer, conduire, communiquer et suivre des évaluations crédibles au moment opportun.

Ces cinq à huit dernières années, certaines améliorations considérables se sont produites en ce qui concerne la pratique de l'évaluation, notamment pour ce qui est de la qualité des évaluations produites et de la mise en place de nouvelles manières d'intégrer l'évaluation au sein de processus organisationnels plus larges (PAM, 2009 ; Jones *et al.*, 2009 ; ACF, 2011b ; UNICEF, 2013 ; Tearfund, 2013). Néanmoins, nombreux sont ceux dans le secteur qui ont le sentiment que le plein bénéfice du potentiel des évaluations humanitaires n'est pas atteint actuellement et qu'elles pourraient être mieux ancrées dans la culture, les processus et la structure des organisations humanitaires. Le travail de recherche d'ALNAP confirme cela : les opportunités de maximisation des bénéfices tirés des évaluations ne sont pas toujours saisies et il existe des obstacles au niveau de l'utilisation des résultats d'évaluation qui doivent être surmontés (Sandison, 2006).

Une étude récente portant sur l'état des pratiques humanitaires au sein des organisations caritatives et à but non lucratif aux États-Unis a conclu que plus de deux tiers des organisations sur lesquelles portait l'enquête « n'ont pas les capacités et comportements en place pour s'engager de façon significative dans l'évaluation ». La même étude a observé que l'évaluation est souvent un outil sous-estimé et négligé pour améliorer les résultats et maximiser l'impact : « quelque chose qu'il est bon d'avoir, mais pas forcément quelque chose de nécessaire » (Innovation Network, 2013 : 1). Malgré certains progrès récents au niveau des organisations (à la fois individuellement et entre elles), la précédente observation fait néanmoins écho à certaines réflexions souvent faites en matière d'évaluation pour l'ensemble du secteur humanitaire (voir par exemple : Lindahl, 1998a, 1998b ; Crisp, 2000 ; Griekspoor et Sondorp, 2001 ; Wood *et al.*, 2001 ; Frerks et Hilhorst, 2002 ; Darcy, 2005 ; Fearon, 2005 ; Ridde et Sahibullah, 2005 ; Mebrahtu *et al.*, 2007 ; Guerrero *et al.*, 2013).

On constate des efforts encourageants pour s'attaquer à ces problèmes sur un certain nombre de fronts. Par exemple, le Département du développement international du gouvernement britannique (DFID) a lancé une nouvelle politique d'évaluation soulignant

le rôle de l'évaluation durant la phase de conception de l'intervention (DFID, 2013 : 2, paragraphe 10). Le DFID a également commandité des études analysant les méthodes d'évaluation (Stern, 2012 ; Vogel, 2012 ; Perrin, 2013) et explorant comment améliorer l'apprentissage à partir de l'évaluation et de la recherche (Jones *et al.*, 2009 ; Jones et Mendizabal, 2010). L'UNICEF, suite à un examen de son fonctionnement pour l'évaluation humanitaire (Stoddard, 2005), a travaillé à renforcer la qualité interne de ses produits d'évaluation (UNICEF, 2010a) tandis que le bureau des évaluations du Programme alimentaire mondial (PAM), suite à une revue entre pairs du Groupe d'évaluation des Nations unies (UNEG), a commandité certains travaux visant à améliorer l'objectif d'apprentissage des évaluations de l'organisation (PAM, 2009).

Ces cinq dernières années ont marqué une hausse remarquable du nombre d'évaluations en temps réel (ETR) inter-organisations suite à des crises majeures comme le cyclone Nargis (IASC, 2008), les inondations au Pakistan (IASC, 2011a) et le séisme d'Haïti (IASC, 2011b). Cela s'est accompagné du développement de procédures ETR validées entre organisations (IASC, 2010). Le projet Emergency Capacity Building (ECB) a développé des outils d'orientation sur les questions d'évaluation conjointe (2011) et a conduit une série d'ateliers au niveau national et régional pour renforcer les capacités des membres d'ECB à commanditer et réaliser des études d'impact (ECB-AIM Standing Team, 2011 ; ECB, 2007). Plus récemment, Oxfam a développé des outils d'orientation en matière d'évaluation autour de la question de la contribution au changement et des impacts des programmes humanitaires utilisant une approche fondée sur les moyens d'existence (Few *et al.*, 2013).

De plus, une récente résolution de l'Assemblée générale des Nations unies a souligné l'importance d'un fonctionnement efficace des évaluations pour les agences qui font partie de ce système ainsi que du fait de promouvoir la culture de l'évaluation. La résolution exige que les membres du système de l'aide des Nations unies s'assurent de l'utilisation efficace des résultats, recommandations et leçons tirées (Assemblée générale des Nations unies, 2012). Il est également à noter que l'UNEG et le groupe de l'OCDE-CAD ont travaillé à l'amélioration d'approches permettant de réaliser des revues entre pairs des systèmes et structures d'évaluation de leurs membres (Liverani et Lundgren, 2007 ; UNEG, 2011 ; 2012).



Approche de l'étude et remarque méthodologique

Un grand nombre des défis et complexités rencontrés par l'action humanitaire – en particulier ceux liés à l'échelle et au caractère interconnecté de son ambition, mais aussi aux difficultés de compréhension et d'attribution des résultats et des impacts – se retrouvent dans les défis rencontrés par le travail d'évaluation. Cependant, le sentiment que le plein potentiel de l'évaluation à apporter des changements dans la pratique humanitaire n'est pas atteint actuellement ne s'explique pas seulement par la complexité du système. Il existe de nombreuses autres raisons qui sont le sujet même de cet ouvrage. Cette discussion est enrichie de suggestions de solutions possibles utilisées par les organisations humanitaires pour relever ces défis.

La publication de cette étude représente la conclusion de la deuxième et dernière phase d'une série de travaux menés par ALNAP pour analyser les questions relatives aux capacités de l'évaluation. Comme noté précédemment, le déclencheur de cette étude fut la publication en 2006 d'une étude portant sur l'utilisation de l'évaluation dans le secteur humanitaire. Cet ouvrage suggérait que d'autres études seraient nécessaires pour améliorer la compréhension du fonctionnement et de l'utilisation des évaluations dans la pratique humanitaire (Sandison, 2006 : 90, 139–141).

S'appuyant sur l'intérêt démontré par les membres d'ALNAP pour davantage explorer ces sujets, ALNAP a commandité un travail de suivi pour mieux délimiter la question de recherche et obtenir une idée plus précise des sujets clés de la discussion auprès de ses membres. Cette première phase de la recherche a donné lieu à un article (Hallam, 2011) qui s'appuyait sur des idées développées au moyen d'une revue documentaire, d'entretiens avec des évaluateurs humanitaires et des utilisateurs d'évaluations ainsi que d'un atelier organisé en septembre 2010 avec des évaluateurs des Nations unies, des ONG, du mouvement Croix-Rouge et des bailleurs mais aussi des consultants indépendants et des universitaires. L'atelier a également été utilisé pour consulter les membres d'ALNAP afin de : (i) préciser le sujet de la recherche à venir ; et (ii) explorer quels éléments peuvent faciliter ou freiner les capacités des individus, des équipes et des organisations à financer, commanditer, réaliser et utiliser les évaluations humanitaires de façon plus efficace.

Ce premier article proposait un cadre provisoire des capacités relatives à une évaluation (Hallam, 2011 : 5–6) qui a été conçu pour aider les membres d'ALNAP à renforcer leurs

démarches liées à l'évaluation en les aidant à identifier des domaines prioritaires, à partager des idées entre membres sur ce qui a fonctionné ou non, et à développer de nouvelles stratégies pour s'attaquer aux questions récurrentes. Sous sa forme provisoire, ce cadre visait également à fournir un canevas sur lequel placer les éléments collectés entre 2011 et 2013 au moyen de consultations de groupes et d'entretiens individuels avec des responsables d'évaluation, des évaluateurs humanitaires et du personnel de programme travaillant en lien avec la commande ou l'utilisation d'évaluations¹.

La phase 2 de cette recherche a impliqué une série d'ateliers organisés à Londres, Genève, Washington, Madrid et Tokyo, durant lesquels le premier article a été débattu. Progressivement, une compréhension plus nuancée des défis et des opportunités relatifs aux capacités de l'évaluation a ainsi émergé. Les idées développées lors de ces réunions s'attachaient à promouvoir l'utilisation de l'évaluation au sens le plus large et à améliorer les capacités d'évaluation. Pour la version finale de cette étude, des entretiens supplémentaires ont été menés ainsi qu'une nouvelle recherche portant sur les documents publiés ou non (littérature grise) en matière d'EAH et d'évaluation des actions de développement. ALNAP a également conçu et lancé une communauté de pratique (CdP Évaluation d'ALNAP) dont l'objectif est de solliciter et de recueillir des suggestions, des commentaires et des récits d'expérience de la part des évaluateurs membre du réseau (voir Encadré 1).

Depuis sa phase de démarrage, cette recherche a été collaborative et a reposé sur un mélange de revue documentaire et d'activités plus consultatives et participatives. Cette combinaison a permis aux évaluateurs de différentes organisations du réseau d'exploiter des exemples concrets et, nous l'espérons, certaines histoires inédites d'évaluation ainsi que certains regards d'évaluateurs qui sont rarement documentés dans des articles plus formels et des rapports d'évaluation. Une liste des contributeurs est disponible à la page 90

En résumé, faisant suite au premier article de recherche sur le même sujet (Hallam, 2011), cette étude est le résultat de :

- a. un second tour de consultation de groupes auprès d'évaluateurs ;
- b. un second examen de la littérature portant sur l'évaluation à la fois propre au

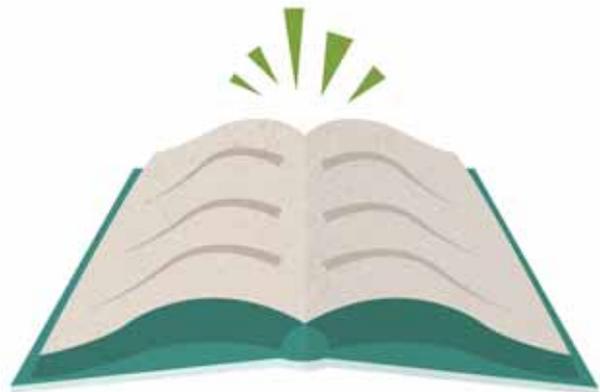
.....
¹Entre 2010 et 2013, ALNAP a animé sept ateliers dans cinq pays. Ceux-ci ont attiré plus de 115 participants provenant de plus de 40 organisations. Les documents de ces ateliers sont disponibles sur : <http://www.alnap.org/using-evaluation>

secteur humanitaire, mais aussi plus largement sur l'évaluation dans le secteur du développement et d'autres secteurs ;

- c. des entretiens individuels avec différents personnels d'organisations plus ou moins impliqués dans des travaux d'évaluation et correspondant aux différents profils représentés au sein d'ALNAP : agences des Nations unies, Mouvement international de la Croix-Rouge et du Croissant-Rouge (CR/CR), bailleurs de fonds, ONG, consultants indépendants et universitaires ;
- d. une conversation à distance avec des membres d'ALNAP durant une année (juin 2012 à mai 2013) sur la CdP Evaluation où les points de vue de différents évaluateurs ont été discutés, développés, réfutés ou validés par différents contributeurs.

Il semble prometteur que de nombreuses organisations et agences accordent à présent une bien plus grande attention aux capacités d'évaluation et à l'utilisation de l'évaluation dans son sens le plus large. Leur démarche et leurs idées ont énormément enrichi ce document.

Certaines histoires inédites d'évaluation ainsi que certains regards d'évaluateurs sont rarement documentés dans des articles plus formels et des rapports d'évaluation.



Encadré

1

Une communauté de pratique pour la consultation et l'apprentissage entre évaluateurs

En juin 2012, ALNAP a mis en place une communauté de pratique de l'Évaluation (CdP) à la fois présentielle et virtuelle dotée d'un double objectif:

1. Faciliter une consultation des évaluateurs du réseau sur la pertinence, la clarté et l'utilité d'un cadre provisoire – proposé dans un précédent document d'ALNAP (Hallam, 2011) – portant sur les capacités d'évaluation humanitaire.
2. Offrir une opportunité pour que les professionnels de l'évaluation réunis au sein d'ALNAP se réunissent – en face-à-face ou en ligne – pour poser des questions, partager leur apprentissage et obtenir des conseils pratiques de collègues et de pairs confrontés à des situations similaires.

Les ateliers ont été accueillis par plusieurs membres d'ALNAP pour discuter de sujets liés à l'appui organisationnel en matière d'évaluation. Les questions de leadership, de culture, de communication, de diffusion et d'adoption de l'évaluation ont été particulièrement ciblées. Ces réunions d'apprentissage entre pairs ont attiré des responsables de gestion, de planification et de coordination, des chargés de relations avec les bailleurs et des conseillers stratégiques ainsi que des évaluateurs^[2].

² Les comptes-rendus des ateliers et autres documents issus des consultations et des événements entre pairs sont disponibles sur : www.alnap.org/using-evaluation.

Objectifs de l'étude

Le principal objectif de ce document est de motiver et d'encourager les évaluateurs humanitaires en soulignant et en discutant des réponses concrètes au défi que constitue l'utilisation insuffisante ou inefficace de l'évaluation. L'utilisation des points de vue et des études de cas provenant de collègues évaluateurs dans le secteur humanitaire est conçue pour renforcer le sentiment que beaucoup de choses sont possibles. Nous espérons que cette étude contribuera à renforcer les capacités des individus, des équipes et des organisations afin qu'ils ou elles deviennent des commanditaires, des utilisateurs et des consommateurs d'informations évaluatives et de connaissances mieux informés. Pour cela, les trois principales approches sont :

- Souligner les sujets clés liés à l'utilisation de l'évaluation dans le secteur humanitaire ;
- Présenter un cadre d'analyse et certaines suggestions pratiques pour améliorer les capacités des organisations à financer, commanditer, soutenir, réaliser et utiliser de façon significative les évaluations humanitaires ;

- Approfondir les sujets en fournissant des études de cas et des points de vue d'évaluateurs qui explorent ce qui a fonctionné dans la pratique.

Le principal public de ce rapport est le personnel des organisations humanitaires. Il existe de nombreuses publications sur le renforcement des capacités dans le monde du développement mais très peu qui s'adressent directement au public humanitaire. Cet ouvrage cherche à combler cette lacune. En cette période de crise économique, la question n'est pas tant de dépenser moins et de faire plus mais de se concentrer sur les stratégies que toutes les personnes impliquées dans le travail d'évaluation doivent être capables de comprendre et sur lesquelles elles doivent s'appuyer pour augmenter l'efficacité de leur travail : depuis les plus fondamentales jusqu'aux réussites les plus « faciles ».

Au moment de rassembler les études de cas et les points de vue d'évaluateurs, il s'est souvent avéré plus difficile que prévu de les intégrer dans une seule partie du cadre. Au final, et comme cela aurait dû être escompté, nous avons constaté qu'une bonne évaluation consiste à bien faire de nombreuses choses plutôt qu'à nécessairement exceller dans un seul domaine du cadre.

Un cadre pour exploiter au mieux les capacités de l'évaluation des actions humanitaires

Entreprendre un travail d'évaluation et assurer sa qualité ne valent la peine que si l'activité donne lieu à une réelle utilisation des résultats et contribue à améliorer les connaissances parmi les personnes les mieux placées pour les utiliser et apporter du changement et des améliorations dans la pratique (Weiss, 1990 ; Feinstein, 2002 ; Grasso, 2003 ; Tavistock Institute, 2003 : 79 ; Preskill et Boyle, 2008 ; Rist *et al.*, 2013). Si nous souhaitons garantir que les évaluations participent au changement et à l'amélioration, il importe d'explorer comment ces changements sont censés se produire. On part souvent du principe que les évaluations produisent de l'information ou fournissent des leçons qui retournent directement dans le cycle stratégique et sont par conséquent intégrées dans la planification des futurs programmes et projets. De cette manière, un processus d'apprentissage continu devrait exister et mener à une performance sans cesse améliorée (Preskill et Torres, 1999 ; Frerks et Hilhorst, 2002 ; Ramalingam, 2011). Cette compréhension présuppose un modèle de planification rationnel et scientifique. Or, dans une société plurielle, complexe et en proie à des turbulences, les décisions liées aux objectifs et aux programmes sont souvent des compromis politiques qui ne correspondent pas nécessairement aux résultats

des évaluations (Weiss, 1990 ; Pawson et Tilley, 1997 ; Frerks et Hilhorst, 2002 ; Start et Hovland, 2004 ; Rist *et al.*, 2013). L'Encadré 2 décrit différentes manières de mettre en pratique les évaluations.

Encadré

2

Les différentes utilisations de l'évaluation

Un précédent article d'ALNAP portant sur l'utilisation des évaluations (Sandison, 2006: 92-97) a souligné les forces et les faiblesses des organisations humanitaires pour ce qui est d'aller dans le sens de cette utilisation et d'une approche focalisée sur l'utilisateur. L'étude analysait les différentes façons d'utiliser ou de mal utiliser une évaluation. Elle insistait sur le débat classique afférent à l'utilisation de l'évaluation en s'appuyant sur la littérature des évaluations de programmes de la façon suivante :

- **L'utilisation directe ou instrumentale** : des évaluations individuelles peuvent être utilisées directement ou de façon « instrumentale » là où les résultats, les observations, les conclusions et les recommandations sont adoptés. Dans la pratique, c'est peu fréquent et là où cela se produit, cela tend à être fait de manière seulement partielle.
- **L'utilisation conceptuelle** : il est plus fréquent que plusieurs évaluations ou des évaluations individuelles combinées avec d'autres éléments d'information et d'autres points de vue soient utilisées de façon cumulative pour alimenter des débats et influencer des prises de décision.
- **L'utilisation du processus** : même lorsque les résultats d'une évaluation ne sont pas directement utilisés, le processus qui consiste à réaliser une évaluation peut offrir des opportunités pour échanger des informations et clarifier des réflexions. L'utilisation du processus se produit lorsque les personnes impliquées apprennent du processus lui-même plutôt que seulement des résultats de l'évaluation (Patton, 2011: 145).
- **L'utilisation symbolique** : les évaluations peuvent être utilisées pour justifier des préférences et des actions préexistantes. De nombreuses conclusions d'évaluations affirment ce que des personnes ressentent déjà en ce qui concerne les réussites et les manques d'un programme. Une telle utilisation des connaissances n'est pas dramatique, ni parfois même visible, mais elle renforce la confiance des personnes qui veulent encourager le changement (Weiss, 1998: 317) ou justifier des décisions prises par le passé ou encore les utiliser pour du plaidoyer et du marketing (McNulty, 2012).

Le changement n'est pas un processus rationnel mais quelque chose qui se produit à travers diverses interactions complexes à l'intérieur d'une organisation. Cela a pour conséquence que le fait d'exploiter l'ensemble des capacités des EAH en faveur du changement nécessite d'accorder de l'attention à toute une palette de sujets interconnectés. Ces derniers impliquent bien plus que le fait de simplement mieux « utiliser » les évaluations bien que cela soit évidemment souhaitable. Une « réflexion évaluative » doit être récompensée et imprégner l'ensemble de l'organisation. Pour autant, les personnes doivent souhaiter le changement et en être félicitées si l'on veut que des améliorations se produisent dans les programmes. Elles doivent avoir les instruments et le temps leur permettant d'analyser ce qu'elles font et d'explorer les résultats de manière à encourager le changement. L'organisation dans son ensemble a besoin d'accepter que cela va questionner les méthodes de travail existantes et le débat critique. Le leadership et la culture de l'organisation sont d'une importance capitale de même que les systèmes formels et informels, et les mécanismes de gratification.

Pour faciliter l'analyse et la discussion des facteurs qui influencent l'impact des évaluations, nous avons regroupé nos observations dans un cadre d'analyse. L'objectif de ce dernier est d'aider à préciser les discussions ainsi que de stimuler la pensée et l'action en matière d'amélioration de l'impact de l'EHA. Ce cadre s'appuie sur la littérature spécialisée dans le développement des capacités de l'évaluation et l'utilisation des évaluations. Il a été influencé par des discussions avec des évaluateurs humanitaires et par des conversations au sein de la communauté de pratique (CdP) de l'évaluation d'ALNAP (Encadré 1). Ce cadre a été conçu pour être aussi léger et facile à utiliser que possible tout en se focalisant clairement sur les questions les plus pertinentes en matière d'évaluation humanitaire.

Nous présentons ici les grandes lignes de ce cadre, une description plus détaillée de chaque domaine de capacités se trouvant dans les trois chapitres suivants. Les éléments de l'étude de cas et les points de vue d'évaluateurs sont intégrés dans le cadre à l'endroit approprié. Nous observons tout d'abord les questions générales de leadership, de culture, de structure et de ressources (chapitre 1), puis celles d'objectif, de demande et de stratégie (chapitre 2), et enfin, celles plus spécifiques des processus et des systèmes d'évaluation (chapitre 3).



Domaine de capacités 1 : Leadership, culture, structure et ressources

- Garantir que le *leadership* soutient l'évaluation et le suivi
- Promouvoir une culture de l'évaluation :
 - Réduire le risque que l'évaluation soit perçue comme une critique et les évaluateurs comme des critiques
 - Utiliser du personnel à la fois interne et externe pour réaliser des évaluations
 - Renommer les évaluations
 - Être flexible et prendre du plaisir à ce que l'on fait
 - Trouver les bonnes incitations
- Créer des structures organisationnelles qui promeuvent l'évaluation
- Sécuriser des ressources adéquates (financières et humaines)

Domaine de capacités 2 : Objectif, demande et stratégie

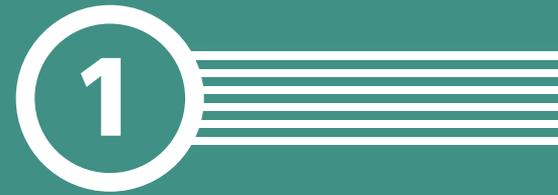
- Clarifier l'objectif de l'évaluation (redevabilité, audit, apprentissage) et l'articuler avec la stratégie générale en matière d'évaluation.
- Augmenter la demande en matière d'informations sur l'évaluation :
 - Rechercher une implication des parties prenantes ;
 - Garantir que les processus d'évaluation se déroulent au bon moment et font partie intégrante du cycle de prise de décision.
- Développer une approche stratégique pour sélectionner ce qui devrait être évalué.

Domaine de capacités 3 : Les processus et systèmes d'évaluation

- Renforcer les processus préévaluation.
- Améliorer la qualité des évaluations :
 - limiter le périmètre de l'évaluation ;
 - impliquer les bénéficiaires ;
 - assurance qualité ;
 - s'engager dans une revue entre pairs de la fonction d'évaluation.
- Diffuser les résultats de façon efficace.
- Renforcer les processus de suivi et post-évaluation, y compris les relations entre l'évaluation et la plus large gestion des connaissances :
 - Garantir qu'il existe une réponse de la direction suite aux évaluations ;
 - Garantir un accès et une facilité de recherche vis-à-vis de l'information évaluative.
- Conduire des méta-évaluations, des synthèses d'évaluation, des méta-analyses et des bilans de recommandations.



DOMAINE DE CAPACITÉS



LEADERSHIP, CULTURE, STRUCTURE ET RESSOURCES

Garantir que le leadership soutient l'évaluation et le suivi	19
Promouvoir une culture de l'évaluation	23
<i>Réduire le risque que l'évaluation soit perçue comme une critique et les évaluateurs comme des critiques</i>	26
<i>Utiliser du personnel à la fois interne et externe pour réaliser des évaluations</i>	29
<i>Renommer les évaluations</i>	31
<i>Être flexible et prendre du plaisir à ce que l'on fait</i>	33
<i>Trouver les bonnes incitations</i>	35
Créer des structures organisationnelles qui promeuvent l'évaluation	38
Sécuriser les ressources adéquates (financières et humaines)	40

Garantir que le leadership soutient l'évaluation et le suivi

Les entretiens réalisés dans le cadre de cette étude ont identifié le leadership comme un élément clé dans l'amélioration de l'impact de l'EAH. On retrouve le même résultat à la fois dans les ouvrages universitaires (Weiss, 1990 ; Torres, 1991 ; Mayne et Rist, 2006 ; Heider, 2011) et dans les articles portant sur les fonctions Évaluation des organisations humanitaires (Foresti, 2007 ; UNEG, 2012a). Les personnes interrogées ont commenté leurs expériences en matière de changement pour ce qui est de l'impact profond et positif du leadership sur la valeur et l'efficacité des évaluations. Là où les dirigeants ne s'intéressent pas à l'évaluation ou sont trop sur la défensive lorsqu'il s'agit d'aborder les questions de performance, il peut exister une certaine réticence dans le fait d'accepter les résultats d'évaluations et les contraintes liées à l'apprentissage tiré des leçons de l'expérience. Si les données et les analyses ne sont pas valorisées au niveau de la direction générale, cela peut se retrouver dans l'ensemble de l'organisation et mener à une réticence, ne serait-ce que pour collecter l'information nécessaire au démarrage. Un leadership de qualité est souvent un élément pivot pour encourager un environnement propice à l'évaluation et à son utilisation ultérieure (Encadré 3) (Buchanan-Smith avec Scriven, 2011 ; Heider, 2011 ; Tavistock Institute, 2003 ; Clarke et Ramalingam, 2008 ; Beck, 2004).

Encadré

3

L'importance du *leadership* dans l'utilisation des évaluations

Un leadership organisationnel qui soutient l'évaluation et utilise de façon réfléchie les résultats des évaluations est essentiel au bon fonctionnement d'un bureau interne de l'évaluation. Les dirigeants d'une organisation doivent persuader leur personnel que l'évaluation fait partie de la vie de l'organisation de manière à obtenir leur soutien aux activités d'évaluation et à construire de la confiance dans les résultats de l'évaluation (Boyne et al., 2004). De plus, les dirigeants doivent identifier l'auto-réflexion et l'amélioration des programmes comme des principes de gestion clés soutenus par l'organisation et encourager le personnel à s'engager dans ces activités grâce à une enquête évaluative (Minnett, 1999). Pour que cela se produise, l'ensemble de l'organisation doit avoir des connaissances en matière d'évaluation (Boyne et al., 2004) et se sentir concernée par le processus de l'évaluation (Bourgeois et al., 2011: 231).



Que peuvent faire les unités d'évaluation pour améliorer le soutien des dirigeants en matière d'évaluation et de suivi ? Voici quelques suggestions qui sont illustrées plus en détail dans des points de vue et des mini-études de cas tout au long de ce document.

- Identifier un « Monsieur Evaluation » qui peut montrer à l'ensemble de l'organisation les bénéfices pouvant être tirés d'une évaluation.
- Attirer l'attention sur des évaluations de bonne qualité qui soulignent des points importants au sein des programmes et qui auraient pu être oubliés sans elles car cela peut servir d'outil pour convaincre les dirigeants de l'importance de l'évaluation.
- Chercher à s'assurer que la politique de recrutement de l'organisation pour ses hauts dirigeants et que des séances d'orientation soulignent l'importance de l'évaluation et de son utilisation comme outil d'apprentissage et de réflexion.
- Nommer des dirigeants dotés de responsabilités spécifiques en matière d'évaluation et de gestion des connaissances.
- S'assurer que l'unité Évaluation de l'organisation se focalise sur les besoins en informations de la direction générale afin que celle-ci continue de s'engager et de soutenir la démarche.



Regard d'évaluateur 1

Focaliser l'évaluation sur des questions provenant de la direction générale

L'UNRWA a récemment connu une période de transformation de plusieurs années qui a offert à l'unité Évaluation l'opportunité d'identifier des questions d'utilisation particulière pour les dirigeants de l'organisation responsables des prises de décision stratégiques. Les évaluateurs travaillent avec le comité de direction pour identifier des questions pertinentes aux yeux d'unités spécifiques durant une phase particulière du processus et vérifient avec d'autres parties prenantes clés comme les bailleurs et les organisations hôtes de manière à s'assurer que ces derniers pensent aussi ces questions pertinentes. Cela permet de prioriser les évaluations qui permettent à l'organisation de progresser.

Source : Représentant de l'UNRWA, communication personnelle lors d'une réunion de l'UNEG, avril 2013.

Une étude sur le renforcement de l'apprentissage issu de la recherche et de l'évaluation au DFID (Jones et Mendizabal, 2010) a suggéré la nomination d'un directeur au rôle de « Monsieur connaissances et apprentissage ». Dans la même logique, l'UNICEF a créé un poste entièrement dédié à la gestion des connaissances humanitaires et aux capacités d'apprentissage au sein de son unité d'urgence. Cette même organisation a également mené récemment une revue de ses évaluations d'action humanitaire (2013) dont les recommandations sont les suivantes :

- L'orientation de l'EAH doit être comprise comme une séance du programme d'orientation des Représentants et de leurs adjoints d'ici la fin 2013.
- La sensibilisation à l'EAH sera comprise dans les réunions régionales qui commencent en 2013 et se poursuivent au-delà (par exemple, les réunions régionales de gestion, les réunions des représentants adjoints et du personnel des Opérations ; les réunions régionales d'urgence).



Regard d'évaluateur 2

Un changement au sein de la direction, une opportunité pour l'évaluation?

Les unités ou les équipes d'évaluation doivent être capables et désireuses de tirer profit de certaines fenêtres d'opportunité pour catalyser les discussions sur l'évaluation, l'apprentissage et l'amélioration. Un changement au sein de la direction générale peut fournir une telle fenêtre. Par exemple, les faiblesses de la réponse humanitaire d'une agence des Nations unies ont été régulièrement soulignées par des évaluations durant une décennie, mais celles-ci recevaient peu d'attention en termes de suivi. Or, la réémergence de ces défis durant la réponse au tremblement de terre d'Haïti s'est trouvée coïncider avec un changement au niveau de la direction générale si bien que cela a été l'occasion d'une nouvelle manière de voir les choses.

Le nouveau directeur général était désireux d'apprendre et de s'impliquer dès le tout début dans une évaluation commanditée en Haïti. L'unité Évaluation a utilisé les connaissances préalables pour enrichir les discussions et a mené une analyse concrète qui a été présentée de façon adaptée à la nouvelle direction afin qu'elle agisse sur la base de ces informations. L'engagement de la direction générale a été vital pour que cette étude demeure pertinente par rapport aux principales tendances stratégiques de l'organisation et a permis qu'une attention et une implication politique perdurent vis-à-vis de l'étude. De plus, l'implication du Cabinet du directeur a aidé à garantir une réponse rapide de la direction qui a amené des changements majeurs au niveau de l'amélioration de la performance de l'organisation lors des urgences de grande échelle. Ces changements comprenaient : des procédures d'opérations standards simplifiées, une plus grande prise en compte du travail en Cluster à travers des formations et des instructions pour la direction générale, des processus de ressources humaines affinés pour avoir les bonnes personnes sur le terrain au bon moment, et des instructions renforcées pour les réponses aux catastrophes urbaines.

Source : Abhijit Bhattacharjee, consultant indépendant, communication personnelle, juin 2013.



Regard d'évaluateur 3

Le soutien de la direction vis-à-vis de l'évaluation : un exemple issu de l'évaluation du Cluster Global Education

UNICEF et Save the Children ont réalisé une évaluation du Cluster Global Education. Ce travail a été très apprécié et bien utilisé. Une des principales raisons de ce succès a été que, en raison de l'importance de l'approche Cluster pour ces deux organisations, les deux directions ont bien soutenu l'évaluation même si elles savaient que celle-ci contiendrait des critiques à leur égard. Suite à ce travail, une réponse conjointe des directions a été lancée : les objectifs du Cluster et les indicateurs ont été clarifiés de même que les rôles et responsabilités ; une planification, un budget et une mobilisation des ressources ont été préparés de façon conjointe ; et des accords de gouvernance renforcés ont été conçus et mis en œuvre. L'évaluation a été assez critique vis-à-vis de l'UNICEF dans certaines zones mais cela n'a empêché ni une très bonne réponse de la direction ni un suivi actif après sa conclusion.

Source : Rob McCouch, UNICEF, communication personnelle, octobre 2012.

Promouvoir une culture de l'évaluation

Favoriser une évaluation crédible qui est à la fois utile et utilisée nécessite de mettre en place des systèmes et des structures efficaces qui soutiennent le travail d'évaluation au niveau de la budgétisation et de l'allocation des ressources humaines mais aussi de la gestion des connaissances et des processus d'apprentissage (Rist *et al.*, 2011 ; Heider 2011). Cependant, Patton (2008) et Mayne (2008) ont tous les deux souligné que cela est seulement secondaire par rapport à la nécessité de mettre en place les deux éléments clés que sont la culture et le leadership. Selon Mayne (2008), les efforts pour créer des systèmes d'évaluation sans traiter la question de la culture organisationnelle ont de grandes chances de se solder par un système pesant et potentiellement contre-productif. Là où la culture est propice, une organisation est plus à même de rechercher activement des informations liées à la performance, y compris des données d'évaluations, pour apprendre et changer son comportement.

Travailler à encourager et maintenir une culture de l'évaluation n'est pas une fin en soi mais peut devenir un élément moteur pour permettre la constitution d'une organisation orientée vers l'apprentissage (Guba et Lincoln, 1989 ; Weiss, 1998 ; Preskill et Torres, 1999 ; Mayne et Rist, 2006 ; Mayne, 2008 ; Patton, 2011 ; Heider, 2011). Ce changement est soutenu lorsque la promotion d'une culture de l'évaluation comprend les éléments suivants :

- Stimuler la capacité du personnel à demander, générer et utiliser des produits d'évaluation de haute qualité.
- Créer des espaces pour que le personnel puisse parler de la qualité et de la performance des programmes, y compris des éléments critiques et des échecs.
- Comprendre toute la gamme des objectifs et des utilisations des évaluations.
- Inscire davantage l'évaluation dans le cadre de l'apprentissage et de l'amélioration, et moins dans le cadre de la redevabilité, mandatée et verticale (de haut en bas).
- Stimuler la demande en matière de connaissance évaluative au sein du personnel de l'organisation à différents niveaux (du département des programmes jusqu'au département stratégique).
- Faire en sorte que les programmes et interventions soient planifiés, conçus et suivis d'une manière propice à l'évaluation à venir.
- Reconnaître les limites de l'évaluation et le besoin de combiner les éléments de preuves quantitatifs et qualitatifs pour soutenir les résultats et les conclusions de l'évaluation.

- Reconnaître les différents besoins des différents utilisateurs de l'évaluation.^{3]}

Au niveau le plus large, une culture de l'apprentissage et de la redevabilité (à la fois ascendante en direction des institutions et des financeurs, et horizontale en direction des communautés affectées par les crises) peut renforcer l'environnement propice nécessaire à apporter le changement et des améliorations dans la pratique humanitaire. L'« environnement propice à l'évaluation » a été décrit comme le degré avec lequel l'information sur la performance passée a été recherchée et la mesure dans laquelle il existe une dynamique pour continuellement réfléchir, apprendre, améliorer et tenir responsables les personnes des actions réalisées, des ressources dépensées et des résultats obtenus (Heider, 2011 : 89). Une telle culture se retrouve également dans les normes tacites de comportement, y compris la compréhension de ce qui peut et doit être fait (ou non) pour corriger les défauts et renforcer les éléments positifs (Tavistock Institute, 2003 ; Heider, 2011).

Une culture de l'apprentissage et de la redevabilité peut renforcer l'environnement propice nécessaire à apporter le changement et des améliorations dans la pratique humanitaire.



³Bien sûr, de nombreux points évoqués ci-dessus font écho aux principes de base de la littérature portant sur l'utilisation des évaluations (voir par exemple : Chelimsky, 1997 ; Weiss, 1998 ; Feinstein, 2002 ; Patton, 2008).

Dès lors, comment les organisations s'y prennent-elles pour cultiver une culture interne positive ? S'il est vrai que le succès engendre le succès, alors les évaluations d'excellente qualité qui répondent à de réels besoins en informations doivent aussi améliorer la manière dont le travail d'évaluation est lui-même regardé. Cela peut influencer de façon positive la culture de l'évaluation, ce qui, à son tour, augmente la demande d'évaluations et rend plus probable l'amélioration durable des systèmes et des structures d'évaluation.

S'ensuit un cercle vertueux, et s'attaquer aux problèmes à chaque étape du cadre identifié dans cet ouvrage peut aider à créer ce cercle vertueux. Cependant, chaque organisation a également sa propre culture spécifique et sa manière de faire les choses. Par conséquent, chacune devra analyser sa propre situation en cours, son niveau de préparation face au changement et les défis à relever pour renforcer les compétences en matière d'évaluation. Différentes stratégies seront alors requises pour chaque organisation : celles promouvant le changement peuvent devoir être opportunistes et profiter des « victoires faciles » de manière à ce qu'un impact soit rapidement observé et qu'un soutien soit mobilisé en vue de changer la culture de l'évaluation.

Malgré les différences entre organisations, certains thèmes généraux ont émergé tout au long de cette étude sur la manière dont il est possible d'améliorer une culture de l'évaluation :

- Réduire le risque que l'évaluation soit perçue comme une critique et les évaluateurs comme des critiques
- Utiliser du personnel à la fois interne et externe pour réaliser des évaluations
- Renommer les évaluations
- Être flexible et prendre du plaisir à ce que l'on fait
- Trouver les bonnes incitations.

Nous abordons ces sujets dans les pages suivantes.



Ceux promouvant le changement peuvent devoir être opportunistes et profiter des « victoires faciles » de manière à ce qu'un impact soit rapidement observé.

Réduire le risque que l'évaluation soit perçue comme une critique et les évaluateurs comme des critiques

L'évaluation peut générer une appréhension chez ceux qui s'engagent dans une telle démarche ou y sont soumis. Certains peuvent voir les évaluateurs comme des critiques, ce qui donne lieu à des résistances face à leurs recommandations en faveur du changement (voir par exemple Weiss, 1998 : chapitres 2 et 5). Plusieurs conversations récentes avec des membres d'ALNAP soulignent une diversité d'approches pour changer ces perceptions.



Regard d'évaluateur 4

Oxfam Intermón formule ses critiques de façon moins personnelle

L'organisation a utilisé un rapport d'ALNAP sur les leçons tirées des réponses aux inondations⁴ comme document clé d'arrière-plan pour évaluer son programme de réponse aux inondations au Mozambique. L'équipe d'Intermón Oxfam a été capable de lier nombre des résultats issus de sa propre évaluation à ceux détaillés dans le rapport d'ALNAP. Cela a permis à l'équipe de voir que certains problèmes et défis étaient communs à d'autres réponses et rencontrés par d'autres organisations dans d'autres pays, souvent pour des raisons systémiques. Cela a changé la manière dont les personnes dans l'équipe pays ont perçu l'évaluation et l'ont rendue moins effrayante et plus utile.

Source : Luz Gómez Saavedra, Oxfam Intermón, communication personnelle, mars 2013

⁴Alam, K. (2008) 'Flood disasters: learning from previous relief and recovery operations'. London: ALNAP/Provention consortium (www.alnap.org/pool/files/ALNAP-Provention_flood_lessons.pdf).



Regard d'évaluateur 5

Le Groupe URD a adopté un « processus d'évaluation itérative »

Cela aide à créer des relations plus fortes entre le personnel du programme, les évaluateurs externes et l'organisation. Les évaluateurs acquièrent une meilleure compréhension de l'organisation et de son contexte opérationnel. Cette approche aide également à réduire l'idée que les évaluateurs sont « parachutés ».

Le Groupe URD a conduit trois ou quatre évaluations du même projet sur une période de deux ans. Les dernières évaluations se concentrent largement sur l'identification des progrès réalisés avec les recommandations de la mission précédente ainsi que sur l'identification de nouveaux défis. Cela signifie que les évaluateurs doivent réellement s'approprier les recommandations – ils ne peuvent pas simplement les formuler et s'en aller – puisqu'ils vont revenir étudier leur mise en œuvre et leur impact. Cela amène un dialogue puissant entre l'évaluateur et le personnel du programme qui perdure tout au long du projet. [...] Le coût de cette approche réside dans le fait que l'évaluateur perd un peu de son indépendance (mais pas, espérons-le, de son objectivité) de manière à devenir un agent du changement. Mais [...] les bénéfices en termes d'amélioration – qui est, après tout, le principal objectif – font que cela en vaut la peine.

Source : François Grünewald, Groupe URD, contribution à la CdP Évaluation d'ALNAP, novembre 2012

D'autres organisations soulignent les résultats positifs et les exemples de bonnes pratiques de leurs évaluations. Un excellent exemple de cela provient d'Action contre la faim (ACF). Une partie des nouvelles lignes directrices et politiques d'évaluation d'ACF demande aux évaluateurs d'identifier de façon systématique « un aspect particulier démontrant la force de chaque projet évalué d'ACF » (ACF, 2011). Cette pratique tente de « ne pas présenter une approche de programmes de façon exhaustive mais de mettre en valeur un élément fort dont d'autres professionnels d'ACF et des secteurs humanitaire/développement peuvent s'inspirer et tirer des bénéfices » (ACF, 2013 : 31). Une sélection de « bonnes pratiques » qui respectent le cycle commun à plusieurs niveaux opérationnels (encadré 4), lequel est constitué d'étapes pour valider et élaborer l'aspect retenu, est ensuite publiée dans la revue d'apprentissage annuelle (ACF, 2011a ; 2013).

Encadré

4

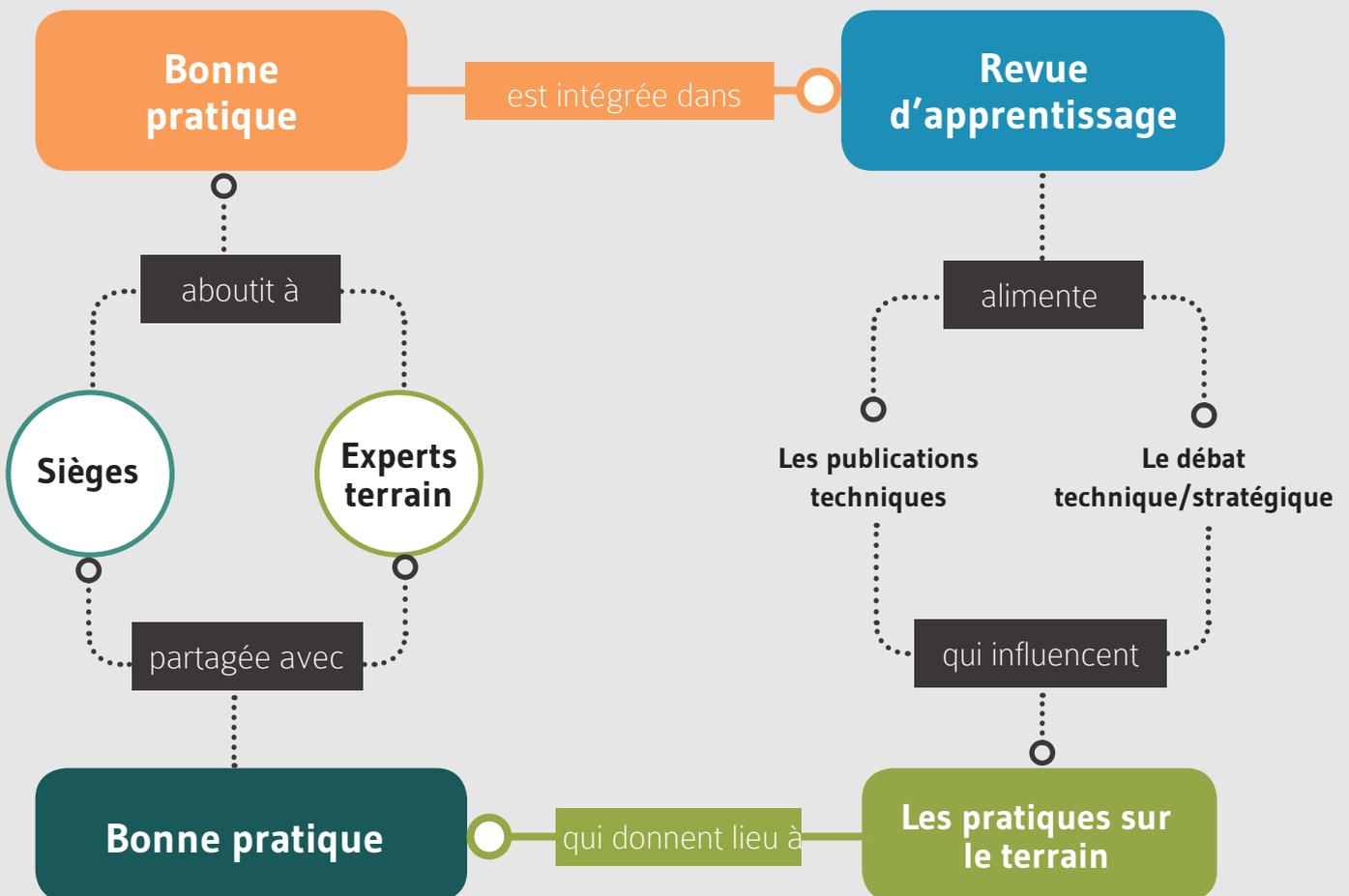
Le processus d'Action contre la faim pour collecter et valider les bonnes pratiques

La pratique récente d'ACF en matière de collecte, d'élaboration et de validation des bonnes pratiques émergentes suit une série d'étapes comme décrit ci-dessous dans la boucle d'apprentissage (à lire depuis le coin en bas à gauche).

De quoi s'agit-il ?

Comme cela fonctionne-t-il ?
Comment avancer ?

Bonnes pratiques de
l'ensemble des 46 missions



Identifiée par les évaluations

Le personnel d'ACF a l'impression que des résultats positifs issus de cet exercice sont déjà observables. Cette pratique a par exemple contribué à améliorer et renforcer les relations avec les évaluateurs car la communication est laissée ouverte – une fois le rapport d'évaluation et les autres produits terminés – et les évaluateurs peuvent voir comment leurs recommandations initiales ont évolué et été adoptées par l'organisation. ACF commence également à voir l'apprentissage devenir un cercle complet à différents niveaux de l'organisation. Par exemple, les « bonnes pratiques » se déploient de façon organique au travers de l'organisation, les équipes basées sur le terrain prenant les commandes de leur adoption. De façon encourageante, l'unité Évaluation et Apprentissage d'ACF a également commencé à noter comment le développement des nouveaux programmes et stratégies fait référence de façon plus explicite aux éléments de preuve évaluatifs qui ont été produits (Guerrero, 2013).

Bien évidemment, la critique ne peut pas toujours être évitée : certaines ONG font le choix d'évaluations internes plutôt qu'externes, jugeant que la critique est plus susceptible d'être acceptée lorsqu'elle vient de personnes de l'intérieur qui comprennent bien l'organisation et qui ont elles-mêmes été impliquées dans la gestion de programmes semblables. Enfin, une explication honnête et franche peut devenir compliquée lorsque les évaluations doivent être rendues accessibles au grand public.

Utiliser du personnel à la fois interne et externe pour réaliser des évaluations

Une revue de la littérature et des discussions avec le personnel de terrain révèle un mécontentement partagé entre le personnel de terrain et le personnel opérationnel quant aux pratiques actuelles d'évaluation. Idéalement, des personnes de l'intérieur seraient nécessaires pour apprendre à connaître la culture et les pratiques de l'organisation qui est en train d'être évaluée. Nombreux sont les employés sur le terrain qui ont l'impression de devoir investir beaucoup de leur temps pour qu'une évaluation amène des recommandations pertinentes et pratiques. Lorsqu'ils ne le font pas, cela peut nuire à la crédibilité de l'évaluation et alimenter la perception souvent entretenue que rien ne change après une évaluation ^[5].

Durant cette étude, de nombreuses personnes interrogées ont indiqué qu'elles ont commencé à utiliser plus d'évaluations internes ou des évaluations mixtes composées de

.....
⁵Pour plus des éléments à ce sujet, voir Patton (1990).
.....

personnes externes et internes à l'organisation. La littérature récente explique que l'un des principaux bénéfices liés à l'utilisation d'évaluateurs internes est leur grande capacité à faciliter l'utilisation d'une évaluation. Par rapport à des consultants externes, des personnes de l'intérieur de l'organisation sont « plus susceptibles d'avoir une meilleure compréhension des préoccupations du personnel sur le terrain et de leurs points de vue sur des sujets clés. Par conséquent, elles sont plus susceptibles de trouver des moyens permettant de soutenir l'évaluation en transformant les résultats en des éléments pouvant être mis en pratique de façon adéquate, et sont plus capables de suivre leurs recommandations au moyen d'activités de suivi » (Bourgeois *et al.*, 2011 : 230). En dehors du fait de soutenir plus fortement une utilisation directe ou instrumentale des résultats de l'évaluation, des évaluateurs internes sont plus à même également d'encourager l'utilisation du processus au sein de l'organisation et de soutenir l'apprentissage organisationnel à un niveau plus élevé^[6].

Malgré ces bénéfices, certains auteurs affirment que des « modèles hybrides » (combinant à la fois du personnel d'évaluation interne et externe) peuvent souvent être le plus approprié (Bourgeois *et al.*, 2011). Certaines organisations insistent sur le fait que des évaluateurs externes maintiennent une indépendance et garantissent que les recommandations radicales ne sont pas supprimées^[7]. Pourtant, en impliquant du personnel interne dans le travail d'évaluation, les unités d'évaluation peuvent également :

1. Jouer un rôle important dans le transfert et la diffusion de connaissances
2. Tirer des enseignements sur les questions les plus pertinentes à traiter dans de prochaines études ou évaluations (Weiss, 1998 : chapitre 2 ; Minnett, 1999)
3. Jouer un rôle dans la préservation de la mémoire organisationnelle et dans le fait d'assurer que des leçons importantes ne sont pas perdues avec le temps (Owen et Lambert, 1995).

.....
⁶Voir également UNEG (2010) et les chapitres 2.2 à 2.5 et 3.9 de Buchanan-Smith et Cosgrave (2013).

⁷Tony Beck et John Cosgrave, consultants indépendants, communications personnelles, respectivement en mai et juin 2013.



Regard d'évaluateur 6

L'expérience de l'UNHCR avec des équipes d'évaluation mixtes

L'unité Évaluation du HCR a expérimenté dans le passé l'utilisation d'équipes d'évaluation totalement indépendantes. Toutefois, cette pratique a récemment changé puisque l'unité Évaluation favorise à présent des équipes plutôt mixtes qui impliquent à la fois des membres de l'unité et des consultants externes. « Les équipes indépendantes peuvent être très chères à sous-traiter, difficiles à gérer et, au final, elles emportent avec elles beaucoup de ce qui a été appris », a noté le chef d'unité qui avançait également que, depuis son poste, « il serait très difficile d'attirer le personnel le plus réputé vers l'unité Évaluation si tout ce qu'il avait à offrir était de gérer des consultants. Attirer des personnes d'autres départements de l'organisation vers l'unité Évaluation signifie que l'expérience et l'approche du terrain sont régulièrement rafraîchies et reflètent l'évolution de la pratique des opérations UNHCR sur le terrain. »

Source : Jeff Crisp, UNHCR, communication personnelle, avril 2012.

CARE est une autre organisation qui utilise des équipes d'évaluation mixtes :

Une évaluation de la tempête tropicale Jeanne a été particulièrement efficace en partie parce que, en plus des consultants externes, le référent suivi et évaluation de CARE Haïti a également rejoint l'équipe d'évaluation, aidant ainsi à garantir que les recommandations seraient réalistes et suivies. Cet employé de CARE a ainsi beaucoup appris. Le rapport a été traduit en français, ce qui a amélioré son utilisation au sein du personnel local et des organisations partenaires. (Oliver, 2007 : 14)

Renommer les évaluations

Mettre de côté le temps nécessaire à l'apprentissage et à la réflexion peut être très difficile à tous les niveaux au sein d'une organisation humanitaire très occupée. Cependant, l'évaluation fait face à des défis bien spécifiques, certains concernant la manière dont elle est aujourd'hui regardée, peut-être en raison de ses mauvaises utilisations dans le passé. (Par définition, une fois que quelque chose est appelée « évaluation », plus personne n'en veut !). Il peut être utile de renommer l'évaluation comme élément d'un changement dans la manière dont les évaluations sont conduites et utilisées. Dans cette logique, certaines personnes interrogées sont allées suffisamment loin pour suggérer que le simple terme « évaluateur » devrait être remplacé par celui de « facilitateur »^[8].

⁸Cette approche fait écho à certaines des réflexions les plus récentes que reflète l'approche de l'évaluation de développement (Patton, 2011).

Là où l'évaluation est utilisée pour un apprentissage et une gestion adaptative, cela implique un changement fondamental du rôle de l'évaluateur. De sa position de personne distanciée, orientée vers la recherche, essayant de systématiser ce qui est connu et de découvrir ce qui est caché, il ou elle devient un facilitateur de processus. Son rôle est d'aider à concevoir et à organiser de façon efficace les recherches et l'apprentissage des autres. L'analyse des parties prenantes et les compétences communicationnelles, de même que la capacité de gérer des dynamiques de groupes, deviennent des atouts majeurs. [...] Renforcer l'apprentissage signifiera au final une profonde restructuration de la formation, de l'approche méthodologique, des compétences professionnelles et de la vision des évaluateurs. (Engel et al., 2003 : 5)

En outre, certaines organisations utilisent l'évaluation comme moyen de promotion et de célébration de leur travail. Le cadre de l'évaluation du Secrétariat du FICR indique par exemple que :

Des évaluations fiables peuvent être utilisées pour la mobilisation de ressources et le plaidoyer ainsi que pour reconnaître et célébrer nos réalisations. La promotion d'un programme ou d'une politique au moyen d'une évaluation n'est pas vue comme une pure tactique marketing parce que les évaluations fournissent des examens impartiaux et souvent indépendants de notre performance et de nos résultats, conférant de la crédibilité à nos réalisations. Elles aident à démontrer ce que nous obtenons en retour de l'investissement de nos ressources ainsi qu'à célébrer nos efforts importants. (FICR, 2011a : 2)

De plus, comme l'a illustré une évaluation de CARE, le fait de souligner ce que l'on a bien fait, plutôt que de se cantonner sur les points où les efforts de réponse se sont révélés insuffisants, peut dynamiser le moral du personnel et augmenter son efficacité (Oliver, 2007).

De même, une étude de la Banque mondiale portant sur l'utilisation des évaluations d'impact a rapporté :

Il existe toujours une demande pour que des résultats viennent confirmer ce que les personnes veulent entendre. Des préoccupations liées à de potentiels résultats négatifs,

à une mauvaise publicité ou à une gestion impropre des résultats peuvent réduire la demande. Dans le même temps, la sensibilité, la construction de la confiance et des mécanismes innovants peuvent aider à surmonter ces craintes. Par la suite, on peut trouver certains bénéfiques à tirer profit d'opportunités pour présenter de bons résultats, en particulier si cela aide le processus à faire en sorte que les parties prenantes comprennent et apprécient le rôle de l'évaluation d'impact.

(Banque mondiale, 2009 : 68)

Enfin, un certain nombre d'organisations produisent des rapports d'évaluation annuels. Certains – comme les rapports d'évaluation annuels du PAM⁹, ou les rapports de capitalisation d'ACDI (ACDI, 2012) – sont bien présentés et contiennent de riches résumés des résultats clés des évaluations. Cela peut aider à rafraîchir l'image de l'unité Évaluation en réduisant l'idée que les unités d'évaluation travaillent seulement à de longs rapports d'évaluations propres à une situation ou à un thème. Cela peut également aider à rapprocher l'évaluation des autres membres de l'organisation. De plus, ces rapports de synthèse peuvent aider à montrer à la direction et à d'autres personnes ce que l'évaluation a permis de connaître.

Être flexible et prendre du plaisir à ce que l'on fait

Une approche pour améliorer la culture de l'évaluation est d'ajouter de la flexibilité, du plaisir et de la créativité au processus d'évaluation. Le directeur de l'unité Évaluation d'une organisation a fait part de son « obsession » personnelle pour ce qui est de se battre avec les rapports d'évaluation ennuyeux :

Les évaluateurs se rendent souvent dans des lieux incroyablement intéressants et rencontrent une grande diversité de personnes hors du commun mais passent ensuite leur temps à écrire des rapports qui sont totalement dénués d'intérêt et de style. Nous avons tous sur nos bureaux bien plus de documents que nous ne pouvons en lire. Des rapports trop longs et mal écrits, pleins de jargon et d'annexes, sont peu susceptibles d'attirer l'attention des parties prenantes clés.

.....
⁹Disponibles à l'adresse suivante : www.PAM.org/about/evaluation/key-documents/annual-evaluation-reports.

Des suggestions pour améliorer les rapports comprennent l'utilisation de citations, d'anecdotes (lorsqu'elles sont pertinentes), de paragraphes courts et de titres percutants pour attirer l'attention du lecteur.



Regard d'évaluateur 7

La flexibilité n'est pas un crime

Il y a beaucoup à dire pour adopter une approche flexible des termes de référence (TdR), les amender comme nécessaire au fil d'un projet d'évaluation, prendre en compte des sujets nouveaux et différents non anticipés lorsque le projet d'évaluation a été lancé. Si quelque chose d'important ou d'intéressant survient au cours d'une évaluation, il serait alors dénué de sens de ne pas réviser en fonction les termes de référence. Ces derniers ne doivent pas être structurés autour des critères classiques de l'évaluation (efficacité, efficience, couverture, cohérence, etc.). Alors que de tels concepts peuvent être utilisés dans un rapport d'évaluation, l'« art » de l'évaluation requiert que l'on évite des approches aussi conventionnelles.

Source : Jeff Crisp, UNHCR, contribution à la CdP Evaluation d'ALNAP, juillet 2012

Une étude de CARE sur l'utilisation de l'évaluation a montré que « le sentiment qui domine largement à propos des rapports d'évaluation est le fait qu'ils sont trop longs et trop fastidieux pour être examinés en profondeur. » (Oliver, 2007 : 2-3).



Regard d'évaluateur 8

La flexibilité en matière de conception d'évaluation

Je crois que la flexibilité dans la conception et la conduite d'évaluations a été notre facteur clé de réussite. Parfois, de courts exercices de capitalisation ont été plus utiles que de vastes processus d'évaluation. À certains moments, il a été enrichissant de prendre du temps pour une évaluation transversale/thématique et réfléchir à des résultats (potentiellement) surprenants grâce à plusieurs débriefings intermédiaires avec des parties prenantes. Nous avons également appris à ajuster le timing des évaluations de projet par rapport au cycle de planification annuelle grâce auquel nous nous assurons que les résultats sont directement intégrés dans le prochain processus de planification.

Source : Sabine Kampmueller, MSF, contribution à la CdP Évaluation d'ALNAP, octobre 2012.

Trouver les bonnes incitations

La synthèse réalisée en 2009 par USAID à propos des tendances en matière d'évaluation conclut que le lien entre l'évaluation et l'action est fonction de la culture et des incitations de l'organisation :

Les mécanismes favorables au traitement systématique des évaluations, plutôt qu'au fait de ranger sur une étagère celles qui donnent des résultats gênants, sont un élément critique des cultures d'évaluation susceptibles d'agir sur les résultats de l'évaluation... Les problèmes surviennent lorsque les incitations au sein des organisations ne sont pas cohérentes avec un modèle d'apprentissage et lorsque l'engagement du personnel vis-à-vis de l'évaluation est inégal en fonction des niveaux hiérarchiques de l'organisation (USAID, 2009 : 27).

Jones *et al.* (2009) ont montré qu'au sein de DFID, le personnel évitait d'intégrer les leçons tirées de projets ou programmes clôturés ou ayant échoué/en train d'échouer parce que les incitations à l'apprentissage étaient en décalage par rapport aux incitations de carrière. A propos des incitations liées à l'évaluation, une récente étude commanditée par InterAction a formulé les suggestions suivantes :

- *Conduire une revue interne des incitations et des mesures désincitatives actuelles en matière d'apprentissage. Une fois connues les manières dont l'apprentissage est découragé (par exemple, comment l'échec est perçu et pénalisé), il est possible de reprendre un processus de construction de la culture de l'évaluation en éliminant toutes les mesures désincitatives.*
- *Commencer à récompenser les réalisations en matière d'apprentissage, – sous la forme de reconnaissance et même de rémunération.*
- *Construire l'utilisation de l'évaluation via les profils de poste et les évaluations de performance. Toute personne qui commande ou contribue à une évaluation doit être tenue pour responsable de ce que les résultats sont utilisés. La « redevabilité » (dans ce contexte) réside dans le fait de créer les conditions favorables à des discussions franches et ouvertes à propos des résultats. La redevabilité sert à faciliter l'apprentissage et non à obtenir de « bons » résultats d'évaluation. (Bonbright, 2012)*

Même les meilleurs plans et les meilleures pratiques sont peu susceptibles de maintenir une utilisation rigoureuse de l'évaluation au fil du temps à moins que des incitations internes et externes ne renforcent l'utilisation efficace des évaluations. Mackay (2009)

observe trois types d'incitations - la carotte, le bâton et les sermons – et rapporte que nombre de ces incitations aident à institutionnaliser le suivi et l'évaluation au sein des gouvernements développés et en voie de développement. Les carottes fournissent un encouragement positif et des récompenses pour la conduite de suivi et évaluation ainsi que l'utilisation des résultats. Les bâtons comprennent des remontrances et des pénalités pour les personnes qui ne prennent pas au sérieux la performance et le suivi-évaluation. Les sermons comprennent les lettres d'engagement de haut niveau et le plaidoyer portant sur l'importance du suivi-évaluation. Bien qu'écrite avec des systèmes de suivi-évaluation nationaux à l'esprit, une sélection des idées les plus pertinentes par rapport à l'EAH est présentée ci-dessous, légèrement adaptée au contexte humanitaire.

Les carottes

- Récompenses ou prix – reconnaissance de la direction vis-à-vis des évaluations réalisées selon les bonnes pratiques.
- Mise à disposition de financements supplémentaires pour la réalisation du suivi-évaluation.
- Assistance aux zones de programme (via des conseils du siège, des formations gratuites, etc.) pour la réalisation du suivi-évaluation. Cela facilite la conduite d'activités de suivi-évaluation et l'utilisation des résultats.

Les bâtons

- Elaborer des politiques qui rendent obligatoires la planification, la conduite et le reporting en matière de suivi-évaluation.
- Retirer une partie du financement si le suivi-évaluation n'est pas fait.
- Pénaliser la non-réalisation des recommandations de l'évaluation qui ont été actées.

Les sermons

- Lettres d'engagement de haut niveau par les directeurs exécutifs et les autres cadres supérieurs.
- Les séminaires/ateliers de sensibilisation visant à démystifier le suivi-évaluation rassurent quant à sa faisabilité et expliquent aux participants ce qu'il recouvre.
- Utiliser des exemples concrets de l'impact du suivi-évaluation pour démontrer son utilité et sa rentabilité.

Même les meilleurs plans et les meilleures pratiques sont peu susceptibles de maintenir une utilisation rigoureuse de l'évaluation au fil du temps à moins que des incitations internes et externes ne renforcent l'utilisation efficace des évaluations.



L'importance de trouver les bonnes incitations est soulignée par Jodi Nelson, directrice de la stratégie, de la mesure et de l'évaluation à la Fondation Bill & Melinda Gates :

Ce que j'ai appris de plus important est que mon travail consiste plus dans du changement organisationnel que dans le fait d'être un expert en évaluation. Une organisation peut avoir un excellent personnel et de l'expertise en matière de suivi-évaluation, comme c'est le cas pour nous, mais cela ne mènera pas forcément à quelque chose à moins qu'il n'y ait un alignement du haut en bas de l'organisation sur la question de ce qui rend possible le succès. On retrouve parmi les exemples : des chefs qui demandent continuellement à leurs équipes de définir et planifier des résultats mesurables, des attentes cohérentes pour le personnel et les partenaires par rapport à ce qui constitue une preuve crédible au niveau de la prise de décision, des dirigeants qui comprennent et financent le changement, lequel peut être difficile et demander beaucoup de temps, mais aussi des outils et des ressources pour que le personnel et les bénéficiaires intègrent une planification et un suivi-évaluation rigoureux dans leur travail quotidien. (Nelson, 2012)

Jones et Mendizabal (2010) ont observé que la demande en éléments de preuve augmentait dans des circonstances où une réponse rapide était rendue nécessaire par un événement inattendu. Dans ces situations, la capacité du personnel à utiliser des éléments de preuve et des leçons tirées de la recherche et des évaluations aide à construire leur crédibilité et

leur influence au sein de DFID. Plus concrètement, les éléments qui ont prouvé le rapport qualité-prix d'un domaine stratégique particulier ont été fortement prisés car cela était utile à la direction de DFID pour renforcer ses arguments en faveur d'une augmentation de ses budgets auprès du Trésor britannique (soulignant l'importance de messages issus du haut de l'organisation).

Cependant, d'une manière générale, Jones et Mendizabal (2010) ont remarqué que des réflexions originales et des idées novatrices étaient plus susceptibles d'être récompensées que l'utilisation d'idées éprouvées ou de leçons apprises ailleurs. Le fait d'orienter les incitations vers des réflexions novatrices, mais compréhensibles, est susceptible de militer contre l'utilisation efficace de l'évaluation. Cette même étude a également observé que les pressions de temps encourageaient les employés à se fier à l'expérience plutôt qu'aux éléments de preuve apportés par la recherche et l'évaluation au moment d'étayer leurs arguments.

Créer des structures organisationnelles qui promeuvent l'évaluation

Les unités d'évaluation ont diverses localisations et lignes de reporting au sein des organisations (Foresti, 2007)¹⁰. Certaines organisations ont une unité Évaluation séparée et/ou centrale responsable de la conduite des évaluations. Dans d'autres organisations, la responsabilité est décentralisée (Hallam, 1998). Le type de structure de gestion influence potentiellement, en bien ou en mal, l'impact de l'EHA. Par exemple, le fait d'avoir une unité centrale consacrée aux évaluations peut améliorer la qualité des évaluations mais, en plaçant l'évaluation dans un « silo », peut également gêner un apprentissage institutionnel plus large.

Certains départements d'évaluation rendent compte directement au Bureau ou au directeur exécutif. D'autres départements d'évaluation sont plus bas en termes de hiérarchie. Une étude commanditée par l'Agence française de développement (Foresti, 2007) a conclu qu'il n'existait pas en la matière de bonne ou mauvaise approche mais, que la position du département avait un impact sur l'apprentissage et la redevabilité aussi bien que sur l'indépendance perçue de l'unité Évaluation. De plus, l'implantation du service Évaluation est susceptible d'avoir une influence sur sa redevabilité vis-à-vis de la direction générale et sur le niveau de résistance qu'elle pourra rencontrer auprès des programmes et du

¹⁰Voir également Webster et Walker (2009).

personnel. La même étude a également remarqué qu'il existait généralement « une tension entre l'indépendance des départements d'évaluation et leur réussite pour ce qui est d'impliquer les utilisateurs de l'évaluation » (Foresti, 2007 : 8). Il est nécessaire de trouver un équilibre entre cette indépendance et d'autres éléments comme la crédibilité et l'utilisation. Trop d'indépendance peut mener à un manque d'utilisation alors qu'une trop faible indépendance peut réduire la crédibilité^[11].



Regard d'évaluateur 9

L'exemple de Tearfund pour intégrer l'évaluation à l'ensemble des services de l'organisation

Le service Suivi-Evaluation de Tearfund est basé au sein de l'équipe d'Appui et Conseil en informations d'apprentissage (qui fait partie du Groupe de développement organisationnel et individuel) et rapporte au directeur de la gestion des connaissances. L'équipe travaille dans toute l'organisation, maintenant de fortes relations de travail avec le groupe international et le chef de la stratégie aussi bien qu'avec les équipes en lien avec les bailleurs ou d'autres partenaires. Cette implantation au sein de Tearfund semble favorable à la fertilisation croisée de l'apprentissage et des idées. Le service Suivi-Évaluation négocie avec l'ensemble des équipes et des groupes, contribuant à hausser le niveau en matière de réalisation et d'utilisation des évaluations.

Source : Catriona Dejean, Tearfund, communication personnelle, juin 2013.

Ce qui fonctionne pour une organisation peut ne pas fonctionner pour une autre. Ainsi, dans une petite ONG, il peut être largement préférable que le bureau Évaluation soit proche des programmes et structurellement moins indépendant. Pour une grande agence des Nations unies où il peut exister d'énormes différences de pouvoir et de perception entre les niveaux de l'organisation, un bureau d'évaluation structurellement plus indépendant peut s'avérer nécessaire pour être en mesure de « dire la vérité à ceux qui détiennent le pouvoir ».

Indépendamment de la taille d'une organisation et de ses mécanismes de reporting, un niveau minimum d'indépendance est nécessaire et l'équipe d'évaluation ne devrait pas être sujette à de la pression et à de la compétition qui peuvent émerger des départements opérationnels. (Heider, 2011) L'implantation du département Évaluation doit être propice à la réduction des barrières en matière d'utilisation de l'évaluation^[12].

¹¹John Cosgrave, consultant indépendant, communication personnelle, juin 2013.

¹²Riccardo Polastro, DARA, communication personnelle, mai 2013.

Il est intéressant de noter que nous n'avons trouvé aucune étude de cas démontrant comment un changement dans la structure d'une organisation avait amené un changement dans l'utilisation et l'impact de l'EAH. Peut-être existe-t-il une lacune au niveau de la recherche ou peut-être cela montre-t-il que, même si l'on s'attend à ce que la structure soit très importante, d'autres éléments (comme la culture et le leadership, ainsi que les processus d'évaluation) le sont encore plus.

Sécuriser les ressources adéquates (financières et humaines)

De nombreux employés et cadres sur le terrain se plaignent des coûts des évaluations. Pourtant, ces personnes ne prennent pas toujours le temps d'envisager les coûts financiers et humains potentiellement immenses qu'implique la poursuite de programmes insatisfaisants. La recherche en matière de coûts et de bénéfices des évaluations pourrait mener à une meilleure compréhension des bénéfices concrets, démontrant que : « Améliorer les évaluations humanitaires via un processus continu de réflexion et d'apprentissage est la manière la plus rentable d'assurer que les programmes humanitaires s'améliorent. » (Apthorpe *et al.*, 2001).

Un rapport analysant l'état de l'évaluation au sein des organisations américaines à but non lucratif (Innovation Network, 2013) a remarqué que les trois barrières les plus significatives en matière d'évaluation étaient :

1. Le manque de temps du personnel
2. Des ressources financières insuffisantes
3. Une expertise limitée du personnel en matière d'évaluation

« Mon travail consiste plus dans du changement organisationnel que dans le fait d'être un expert en évaluation. »



La ressource la plus importante pour les départements d'évaluation est leur personnel. Malheureusement, le personnel des départements d'évaluation n'est pas toujours aussi bien soutenu qu'il devrait l'être. « Les mécanismes incitatifs des organisations ne récompensent pas nécessairement les équipes des départements d'évaluation qui, de ce fait, ne sont pas en mesure d'offrir des opportunités de carrière claires aux membres de leurs équipes » (Foresti, 2007). Un rapport d'organisation de 2007 a remarqué que « les postes [au sein du bureau d'évaluation] sont extrêmement stressants en ce qui concerne les relations entre le personnel et les autres membres de l'organisation... Certains ont eu l'impression qu'une affectation au sein du bureau d'évaluation était une mauvaise évolution de carrière et que cela pouvait affecter la suite de la carrière d'un individu » (Baker *et al.*, 2007).

Étant donné l'existence ou l'absence de tels mécanismes incitatifs, il se peut que les départements d'évaluation ne fonctionnent pas à leur niveau optimal. Le personnel en charge des évaluations doit être formé au niveau approprié et avoir accès à des outils de renforcement des compétences ainsi qu'à des opportunités de formation de recyclage.



Regard d'évaluateur 10

L'initiative de DFID pour augmenter le soutien à l'évaluation

L'étude de synthèse des évaluations stratégiques réalisée par la récente Commission indépendante du Royaume-Uni sur l'impact de l'aide a observé que les compétences limitées du personnel en suivi-évaluation ont été un obstacle majeur à l'appropriation et l'utilisation de l'évaluation. Cela a été davantage souligné par le Bureau national d'audit du Royaume-Uni dans son examen de la gestion de la performance de DFID, lequel a démontré un manque de formation du personnel dans ce domaine. De plus, une évaluation du programme de DFID dans les Caraïbes a conclu que les efforts en suivi-évaluation avaient été gênés par une réduction du personnel qualifié en la matière (Drew, 2011 : 59). Une partie de la réponse de DFID a été de reconnaître l'évaluation comme une profession à part entière, accréditant 133 personnes comme spécialistes de l'évaluation classés de niveau 1 à 4. Il y a également eu une augmentation des opportunités de formation, d'échanges et d'apprentissages en matière d'évaluation ainsi que le lancement de la Conférence annuelle sur le développement professionnel de l'évaluation.

Sources : Government Office for Science, 2011b ; DFID, 2013 ; Jonathan Patrick, DFID, communication personnelle, mai 2012.

Les formations au moment du recrutement et en externe ne sont souvent qu'un seul élément de l'orientation d'un employé en faveur d'un travail en lien avec l'évaluation. D'autres éléments clés concernent des points moins tangibles comme : la création d'une langue commune, avec des références conceptuelles et des idées partagées, ainsi que la construction d'une compréhension commune des objectifs et des pratiques de l'EAH au sein du bureau Évaluation. Nous n'avons trouvé aucune recherche analysant l'impact de la formation à l'évaluation sur la performance des évaluateurs et la qualité de l'évaluation commanditée et réalisée. Certains observateurs ont noté que les opportunités de formation spécifique à l'évaluation de l'action humanitaire étaient relativement rares, et que, tout en étant bénéfique, leur impact se manifeste dans la durée. Il est par conséquent difficile de trouver des exemples concrets prouvant le bénéfice direct de la participation à de telles formations¹³.

En plus d'envoyer du personnel en charge des évaluations à des formations et des cours de recyclage, la rotation du personnel au sein du département d'évaluation a également été encouragée et est régulièrement appliquée dans plusieurs organisations comme l'UNHCR et le PAM. Ce dernier, par exemple, a une politique de rotation de son personnel tous les quatre ans. Au sein de son bureau Évaluation, une moitié du personnel fonctionne sur ce principe de rotation et l'autre moitié est engagée en tant que consultants extérieurs (PAM, 2008b : paragraphe 48).

¹³Margie Buchanan-Smith, consultante indépendante, communication personnelle, juin 2013.

DOMAINE DE CAPACITÉS



OBJECTIF, DEMANDE ET STRATÉGIE

Clarifier l'objectif de l'évaluation (redevabilité, audit, apprentissage) et l'articuler avec les politiques d'évaluation	44
Augmenter la demande en matière d'informations sur l'évaluation	48
<i>Viser l'implication des parties prenantes</i>	50
<i>Garantir que les processus d'évaluation se déroulent au bon moment et font partie intégrante du cycle de prise de décision</i>	53
Développer une approche stratégique pour sélectionner ce qui devrait être évalué	57

Clarifier l'objectif de l'évaluation (redevabilité, audit, apprentissage) et l'articuler avec les politiques d'évaluation

Le fait qu'une évaluation puisse répondre en même temps à des besoins de redevabilité et d'apprentissage est sujet à débat. La littérature spécialisée dans l'évaluation tend à suggérer que ces deux buts sont en conflit, avis que partagent de nombreuses personnes interrogées.



Regard d'évaluateur 11

Être conscient des lignes de faille entre apprentissage et redevabilité

Une session de brainstorming réunissant des évaluateurs humanitaires expérimentés lors d'un atelier d'évaluation a révélé qu'une focalisation trop forte sur la redevabilité (ascendante) dans le cadre d'évaluations était un facteur expliquant leur faible utilisation. Une forte focalisation sur la redevabilité amènent en effet les participants à se sentir jugés et à moins s'impliquer durant le processus d'évaluation mais aussi après, par rapport aux résultats et aux recommandations.

Source : Atelier d'apprentissage entre pairs des évaluateurs membres d'ALNAP, Washington DC, 2011.

Patton note qu'une évaluation commanditée par un bailleur devient souvent une fin en soi à réaliser parce qu'elle est obligatoire et non parce qu'elle est utile. Les évaluations obligatoires peuvent freiner l'utilité en donnant comme raison d'être à l'évaluation la conformité à un besoin de financement plutôt qu'un réel intérêt pour être plus efficace (Patton, 2008).

D'autres observateurs voient de façon moins nette cette ligne de faille entre redevabilité et apprentissage. Tandis que les évaluations se concentrent souvent sur l'un de ces deux objectifs, elles peuvent dans la pratique contribuer aux deux. En effet, certains sont catégoriques sur le fait qu'une évaluation doit répondre aux deux buts et qu'on ne peut réellement apprendre qu'en étant tenu pour responsable. La « vérité » variera selon les organisations et leur culture si bien qu'il n'existe pas ici une seule approche « correcte ». Le Disasters Emergency Committee (DEC) du Royaume-Uni développe par exemple actuellement ses « politique et approche en matière d'apprentissage ». Comme indiqué dans ce document, même lorsque la redevabilité est l'objectif principal d'une évaluation ou



d'une étude que l'organisation a commanditée ou financée : « cela ne doit pas exclure les opportunités d'apprentissage là où elles apparaissent » (DEC, 2013 : 1).

Certaines organisations semblent vraiment peu claires en ce qui concerne l'objectif de leurs évaluations et le fait de reconnaître que différentes approches peuvent être nécessaires pour atteindre différents buts d'évaluation. Certaines avancent que l'amélioration réelle de l'apprentissage et de la performance demande une discussion franche et ouverte au sein d'une organisation, une capacité à partager les échecs autant que les réussites, et enfin une approche dépourvue de jugement. Lorsqu'une évaluation a été commanditée de façon externe et est réalisée par des équipes de consultants extérieurs, une telle approche peut être difficile à atteindre.

L'amélioration réelle de l'apprentissage et de la performance demande une discussion franche et ouverte au sein d'une organisation, une capacité à partager les échecs autant que les réussites, et enfin une approche dépourvue de jugement.



De même qu'aucune évaluation seule ne peut servir les intérêts de tout le monde et répondre aux questions de tous, aucune évaluation ne peut servir seule et avec qualité des objectifs multiples et contradictoires. Des choix doivent être faits et des priorités établies mais au lieu de cela, les évaluateurs, les commanditaires et financeurs d'évaluations, ainsi que les utilisateurs prévus continuent d'attendre que les évaluations servent des objectifs multiples et conflictuels avec la même qualité. Ces attentes non réalistes et l'absence de priorités établies sont souvent ce qui fait que les évaluations commencent à emprunter la longue et périlleuse route qui les mène vers un succès partiel (Patton, 2010b : 151-163).

Une manière de réduire la tension entre les différents buts de l'évaluation consiste à séparer les évaluations orientées vers la redevabilité des évaluations plutôt orientées vers l'apprentissage, et à ne pas essayer de répondre à toutes les attentes avec un seul exercice.

Si les évaluations étaient moins vues comme une menace que comme un « outil d'apprentissage » pour les dirigeants dans le but de les aider à déployer leurs ressources de façon plus efficace, les évaluateurs seraient alors, d'une part, encouragés à écrire des rapports moins prudents et plus utiles, et d'autre part, les dirigeants seraient plus à même d'éprouver de l'intérêt à s'impliquer davantage dans les recommandations faites et leur suivi. (Riddell, 2009 : 13)

Certains soutiennent qu'il serait préférable d'appeler les deux types d'évaluation par des noms différents et de confier la gestion et la réalisation de leurs processus à différents individus et départements au sein de l'organisation. Pour prolonger cet argument, il est possible d'avancer que les évaluations orientées vers la redevabilité (ascendante) pourraient être réalisées par le département d'audit et de redevabilité alors que les évaluations orientées vers l'apprentissage au niveau du terrain et la redevabilité descendante pourraient être réalisées par un département de gestion de la connaissance et de l'apprentissage.

Cette division se produit lentement au sein de certaines organisations. Par exemple, à Sida, il existe aussi bien un comité de surveillance de l'aide qu'une unité interne de l'évaluation, et ce département interne se concentre largement sur l'utilisateur et les évaluations orientées vers l'apprentissage. Différentes organisations ont ajusté leurs allocations de tâches et mis à jour leurs profils de poste pour rechercher l'équilibre entre les différents services Évaluation, Apprentissage, Redevabilité et Audit. Le DEC, par exemple, après une première expérience de recrutement d'un responsable Redevabilité et Audit, et l'allocation des responsabilités liées aux activités d'apprentissage à un autre membre du personnel, a plus récemment fait le choix d'une approche plus conjointe, recrutant un Chargé de redevabilité et d'apprentissage et mettant à jour le descriptif de poste du Directeur des programmes et de la redevabilité pour qu'il soit responsable des deux éléments du reporting du programme (évaluation et apprentissage).

De même qu'aucune évaluation seule ne peut servir les intérêts de tout le monde et répondre aux questions de tous, aucune évaluation ne peut servir seule et avec qualité des objectifs multiples et contradictoires.



Regard d'évaluateur 12

Qui est responsable en matière de gestion et de redevabilité ?

La plupart des commanditaires et contractants d'évaluations continuent de concevoir l'évaluation comme un « outil en faveur de la redevabilité », la redevabilité étant comprise dans le sens classique d'« obligation à rendre compte de la responsabilité conférée ». Voilà qui est totalement faux. La redevabilité fait partie intégrante des responsabilités de gestion, c'est-à-dire qu'il revient à la direction de rendre compte de ses performances et, lorsqu'une tierce partie s'en charge, la direction n'est pas seulement déresponsabilisée, sa gestion est incomplète. Le fait qu'une tierce partie ait la responsabilité de la redevabilité est questionnable, tout du moins dans les prétendus contextes de gestion basée sur les résultats. Et si des individus, à quelque niveau que ce soit, veulent de l'assurance sur leur performance ou des rapports de performance, ils devraient alors recourir à un audit de performance et non à une évaluation.

Source : Ian Davies, XCeval evaluation listserv, janvier 2013.

Le besoin d'évaluation dépend du contexte. La « bonne » approche dépend également de la réceptivité de l'organisation et de sa capacité à tirer des leçons. Une recherche doctorale a demandé aux évaluateurs expérimentés le meilleur type d'évaluation pour différentes organisations. Les évaluateurs ont recommandé que les organisations les moins prêtes pour l'apprentissage s'investissent plus dans la conduite d'évaluations de processus. À l'inverse, celles plus ouvertes au changement devraient plus se focaliser sur des évaluations orientées vers les résultats (Allen, 2010).

Une fois que les choses sont claires, cela peut être reflété dans la politique évaluative d'une organisation qui peut ensuite être communiquée à travers l'ensemble de l'organisation. ACF International a récemment développé une nouvelle politique d'évaluation qui a rassemblé les trois idées clés au cœur de sa vision en matière d'évaluation et d'apprentissage (Guerrero, 2012):

1. Les évaluations n'ont de valeur que si elles sont utilisées.
2. Les évaluations doivent aider à suivre la progression.
3. Les évaluations doivent faire partie intégrante de l'apprentissage organisationnel.

Augmenter la demande en matière d'informations sur l'évaluation

Idéalement, l'« offre » d'évaluation doit correspondre à une demande en produits d'évaluation (encadré 5). Les évaluateurs peuvent prendre plusieurs mesures pour augmenter la demande en informations évaluatives. Le DEC est un exemple d'organisation qui travaille à renforcer ces cultures grâce à une demande en matière de résultats d'évaluation.



Regard d'évaluateur 13

L'intérêt du DEC pour l'apprentissage tiré des réponses antérieures

L'apprentissage est au cœur du fonctionnement du DEC : il constitue l'un des quatre piliers de son cadre de redevabilité et l'une de ses priorités par rapport au fonctionnement de ses membres (DEC, 2013 : 1). Cette dernière composante est bien reflétée dans la manière dont il est demandé aux membres de prendre en compte les leçons apprises dès le tout début du cycle de réponse. Cela signifie que, tout en planifiant leur réponse, les membres doivent prendre en compte quel apprentissage tiré des interventions précédentes sera appliqué. Même si cela n'est pas parfait, car cela peut devenir un exercice de simple vérification ou être transmis au département d'évaluation, cela vise à encourager la réflexion sur l'apprentissage passé et donc l'utilisation.

Source : Annie Devonport, DEC, communication personnelle, juin 2013.

Encadré

5

Quels facteurs influencent la demande des décideurs en faveur de données d'évaluation et de leur utilisation ?

Ouverture : du système vis-à-vis de l'entrée de nouvelles idées et de prises de décision démocratiques.

Un « Monsieur Évaluation » : à un haut niveau politique.

Sensibilisation : compréhension générale au sein des décideurs de l'utilité des données/résultats en matière de suivi-évaluation.

Utilité : l'existence d'un lien entre la prise de décision et le système de suivi-évaluation augmente la perception de l'utilité de l'information.

Incitations : en cas d'utilisation, par exemple en incluant l'utilisation de l'évaluation dans les plans de performance et en reconnaissant publiquement le travail des personnes qui démontrent une bonne utilisation des évaluations.

Influence du bailleur : en particulier, la pression du bailleur sur le pays hôte afin que celui-ci investisse dans l'évaluation.

Source : Levine et Chapoy, 2013 : 5

Une récente étude commanditée par 3ie s'est intéressée à l'implication dans l'évaluation comme possible indicateur permettant de mesurer le niveau de préparation et la volonté des organisations d'aide pour ce qui est de demander de l'information évaluative, mais aussi de commanditer, conduire et utiliser les évaluations. Les auteurs avancent qu'une :

Implication dans l'évaluation est démontrée lorsque les organisations commanditent des évaluations crédibles de façon systématique, rendant publics les rapports et concevant des programmes cohérents avec les recommandations des évaluations.
(Levine et Chapoy, 2013 : 3-4)

En plus de se concentrer sur la création d'une culture réceptive aux évaluations, d'autres phases pratiques existent pour augmenter la demande. Impliquer des parties prenantes (et identifier les principaux utilisateurs prévus comme partie prenante de ce processus) mais aussi s'assurer que des évaluations sont réalisées au bon moment et font partie intégrante du processus de planification en sont deux autres, comme expliqué ci-dessous.

Viser l'implication des parties prenantes

Les ressources étant limitées, il est nécessaire de fixer des priorités pour impliquer les parties prenantes et découvrir des opportunités de grande valeur^[14]. Après avoir identifié vos utilisateurs attendus, vous devez également identifier leurs engagements explicites par rapport à de potentielles utilisations des savoirs issus de l'évaluation^[15]. Or, les utilisateurs prévus sont plus susceptibles de mettre en œuvre les résultats s'ils comprennent le processus et les résultats de l'évaluation, et s'en sentent « propriétaires ». Comme l'étude l'a confirmé, une manière très sûre de créer cette appropriation est d'impliquer activement les utilisateurs dans le processus, et ainsi de construire une relation entre eux et l'équipe d'évaluation ou l'évaluation elle-même. Patton (2008) suggère que l'évaluation débute avec l'élaboration par les utilisateurs finaux des questions auxquelles il faudra répondre.



Regard d'évaluateur 14

Le cycle de planification des évaluations au sein de Sida

Le processus débute par des conversations entre l'unité Évaluation et l'ensemble des unités opérationnelles dans le but de déterminer leurs besoins en connaissances, ce qu'elles veulent savoir et comment l'évaluation pourrait y aider. Cela aboutit à une liste d'environ 100 idées pour des évaluations. À partir de cela, l'unité Évaluation choisit 15 évaluations qu'elle va réaliser. Les évaluations non retenues sont sujettes à des évaluations décentralisées sur lesquelles l'unité Évaluation formule des commentaires et des conseils.

Les employés qui ont proposé l'une des 15 évaluations sélectionnées forment des groupes de référence pour cette évaluation. Ils doivent lister l'utilisation et les utilisateurs attendus. S'il n'y a pas assez d'utilisateurs, l'évaluation est abandonnée. Sinon, ces utilisateurs attendus sont alors impliqués dans la rédaction des termes de référence et travaillent avec le groupe de référence tout au long du processus d'évaluation.

Source : Joakim Molander, Sida, à Hallam, 2011.

¹⁴Bonbright, dans son ouvrage *Use of Impact Evaluation Results*, suggère de poser les questions suivantes :

- Pour cette évaluation, avec ces objectifs, quelles parties prenantes sont les principaux utilisateurs?
- Quel est leur enjeu par rapport à cette évaluation ?
- Quel est leur niveau de pouvoir par rapport aux autres groupes de parties prenantes ?

L'auteur suggère par la suite que des mesures doivent être prises pour soutenir des groupes présentant un niveau de pouvoir plus faible mais qui ont néanmoins un rôle important à jouer dans l'évaluation. Voir également Buchanan-Smith et Cosgrave (2013) aux chapitres 2.5 à 2.7.

¹⁵Pour plus d'informations sur la question de l'engagement des parties prenantes dans une évaluation, voir également Weiss (1983b). Pour des commentaires plus récents, voir Forss et al. (2006) et Oral History Project Team (2006).

La FAO implique de façon systématique les parties prenantes depuis le début des évaluations thématiques ou pays, utilisant des groupes consultatifs qui comprennent des populations affectées, des bailleurs importants et du personnel du programme. Ces groupes s'assurent que les bonnes questions sont posées et ont la possibilité de commenter les versions provisoires des rapports. Cette pratique aide à minimiser l'approche des évaluations qui est souvent définie par l'offre, en particulier celles réalisées dans un objectif de conformité (Encadré 6).

Encadré

6

Qui compose le public de votre évaluation ?

Générer une demande interne peut constituer un défi lorsque les demandes d'évaluations viennent d'en haut. Les chercheurs qui se sont intéressés à l'état de l'évaluation aux États-Unis ont demandé à un échantillon d'ONG d'identifier le principal public de leurs évaluations. Les principaux publics les plus fréquemment cités étaient les bailleurs, suivis par la direction des organisations et d'autres responsables en interne. Les organisations qui s'impliquaient dans des évaluations seulement à la condition d'accéder à des financements étaient peu susceptibles d'attacher de l'importance ou d'avoir un sentiment d'« appropriation » vis-à-vis des produits délivrés.

Source : Reed et Morariu (2010).



Regard d'évaluateur 15

L'expérience du DEC pour construire l'appropriation du travail d'évaluation par les parties prenantes

Le DEC vise à construire de l'appropriation en impliquant ses membres à différentes étapes du processus d'évaluation. Tout d'abord, les TdR de ses Études de réponse (précédemment connues sous le nom d'Évaluations en temps réel du DEC) circulent largement pour des commentaires et des révisions au sein de ses membres. Il est demandé aux membres de partager ces éléments avec les parties prenantes afin de déterminer le « pourquoi » de l'évaluation, c'est-à-dire ce qu'elles aimeraient savoir ou obtenir grâce à cette exercice. Par la suite, une fois sur le terrain, le consultant doit organiser deux réunions avec des représentants de l'organisation et des équipes terrain, que ce soit du personnel ou des partenaires. La première doit clarifier l'objectif de l'organisation et les résultats attendus alors que la seconde doit permettre de restituer les résultats.

Source : Annie Devonport, DEC, communication personnelle, juin 2013.

DFID et CARE ont vu des résultats positifs à partir d'une implication d'utilisateurs même mineure. DFID a observé que l'implication active des personnes interrogées auprès de l'unité Évaluation améliorait leur compréhension du département et aider à réduire les fausses idées quant au fait d'être « audité » (Jones et Mendizabal, 2010). De la même manière, une étude sur l'utilisation de l'évaluation au sein de CARE a remarqué que le fait d'être inclus dans le processus d'évaluation en tant que personne interrogée ou participant à la Revue après action par exemple, augmentait le sentiment d'appropriation des individus par rapport aux recommandations (Oliver, 2007 : 2), les faisait s'intéresser davantage aux résultats et les rendait plus susceptibles de les voir comme « une source d'information légitime pour leur pratique » (Oliver, 2009 : 36–37).



Regard d'évaluateur 16

Tirer le maximum de l'implication du bureau pays dans l'évaluation

Pour optimiser l'utilité au niveau pays, le bureau Évaluation de l'UNICEF a cherché à impliquer le bureau pays à toutes les étapes de l'évaluation grâce à la formation d'un groupe local de référence. Le responsable de l'évaluation est allé au Timor oriental et a impliqué le responsable suivi-évaluation du bureau pays dans les discussions relatives à la stratégie d'échantillonnage et à la conception de l'enquête tout en proposant un appui avec l'assurance qualité de l'évaluation pour dynamiser la qualité générale de l'évaluation.

Cette forme de renforcement des capacités a servi d'incitation au personnel en charge du suivi-évaluation au sein du bureau pays afin d'appuyer activement le programme d'éducation. De ce fait, le personnel s'est également senti plus concerné par son succès. Le résultat a été une évaluation pertinente d'une grande qualité qui a amené une forte réponse de la direction menée par le bureau pays ainsi qu'un processus de suivi. Conséquence de cette approche indépendante et pourtant collaborative, le bailleur s'est engagé dans une seconde enveloppe importante de financement pluriannuel en faveur du programme. Ce processus a également changé la culture du bureau pays qui est devenu bien plus favorable aux évaluations.

Source : UNICEF, 2010b

Patton a avancé lors d'une édition récente du *Canadian Journal of Program Evaluation* (Patton, 2010b) que les réussites partielles et les défauts d'un travail d'évaluation impliquent inévitablement – à un certain niveau et dans une certaine mesure – des relations difficiles et exigeantes avec les parties prenantes. Des évaluateurs expérimentés suggèrent pour cela diverses stratégies : impliquer plus tôt les parties prenantes et s'engager avec elles de façon plus claire, renforcer leurs compétences, gérer les relations avec habileté et traiter les problèmes à temps avant qu'ils ne donnent lieu à des crises.

Garantir que les processus d'évaluation se déroulent au bon moment et font partie intégrante du cycle de prise de décision

Les moments auxquels ont lieu l'évaluation elle-même et la présentation de ses résultats exercent une forte influence sur la demande d'évaluation. De nombreux utilisateurs potentiels de l'évaluation se plaignent qu'elles arrivent souvent trop tard pour être d'une quelconque utilité en matière de prise de décision (Weiss, 1983 ; 1990). Un rapport analysant la recherche et les évaluations de DFID note que :

*La critique la plus fréquente des évaluations au sein des personnes interrogées porte sur la question du timing : bien qu'elles marquent des moments importants pour les bureaux pays, elles sont en général vues comme trop consommatrices de temps par les équipes stratégiques pour être pertinentes. De plus, trop peu de temps est alloué au fait de les relier à des cycles stratégiques et à des fenêtres d'opportunité.
(Jones et Mendizabal, 2010)*



Il n'est pas inhabituel que des évaluations se terminent (ou même débutent) juste après la fin du pic d'une urgence.

C'est encore plus le cas pour les évaluations humanitaires. Un évaluateur senior de MSF remarque que :

Alors que le caractère opportun (timeliness) est un critère particulièrement important dans de nombreuses évaluations, le processus lui-même est souvent extrêmement affecté par la « vitesse d'action » de MSF. Cela signifie que le changement de priorités et le renouvellement des personnes peuvent rendre une évaluation non pertinente alors qu'elle vient juste de se terminer.^[16]

Il n'est pas inhabituel que des évaluations se terminent (ou même débutent) juste après la fin du pic d'une urgence, ce qui présente deux difficultés. Tout d'abord, une analyse importante devient plus difficile à mesure que le temps passe car les souvenirs des informateurs clés s'effacent et des informations se perdent. Ensuite, un long laps de temps entre la fin des programmes et celle de l'évaluation signifie qu'il n'y a pas d'opportunité pour utiliser les résultats afin de changer le programme concerné.

Dans une publication portant sur la manière d'optimiser l'apprentissage issu de l'évaluation, le PAM (2009) a rapporté que le personnel se plaignait fréquemment d'une surcharge d'information, les personnes avançant qu'elles n'avaient pas assez de temps pour lire ne serait-ce qu'une newsletter de deux pages. Dans le même temps, un document de 16 pages produit par ALNAP sur les leçons apprises par les organisations opérationnelles en matière de réponse aux inondations a trouvé de nombreux lecteurs^[17]. D'après une revue documentaire, l'explication provient de ce que l'expérience d'autres organisations et les entretiens apparaissent comme nécessaires pour bien comprendre ce qui est le plus utile à l'utilisateur ciblé et pour trouver le bon équilibre entre contenu, timing et format. La bonne information délivrée à un moment où l'utilisateur la recherche, et sous une forme qu'il peut utiliser, sera saisie au vol. La longueur n'est pas aussi importante que la matière, la forme et le timing.

Comme le dit le chef de l'évaluation de l'UNHCR :

Le timing est un facteur clé pour déterminer l'impact d'une évaluation. Une partie de l'« art » d'un évaluateur consiste à scruter l'horizon, à regarder en avant et à anticiper

¹⁶Sabine Kampmueller, MSF, contribution à la CdP Évaluation d'ALNAP, octobre 2012.

¹⁷Alam, K. (2008) 'Flood disasters: learning from previous relief and recovery operations'. London: ALNAP/Provention consortium (www.alnap.org/pool/files/ALNAP-Provention_flood_lessons.pdf).

ces questions qui seront (devraient être) de la plus haute importance dans le futur proche pour la direction générale. Cela est plus important que le fait que le calendrier des évaluations coïncide avec le cycle de programmation ou de financement.^[18]

De telles préoccupations ont mené au récent développement du nombre d'évaluations en temps réel (ERT) par une diversité d'organisations humanitaires. Une ERT portant sur le séisme en Haïti (Groupe URD, 2010) a été en mesure d'entreprendre une mission terrain moins d'un mois après le tremblement de terre et de saisir les leçons apprises de la réponse immédiate. L'équipe a également été capable de faire une restitution aux acteurs sur le terrain, soit de façon bilatérale, soit aux groupes de Clusters.

Une évaluation en temps réel de la réponse de l'UNICEF à la crise nutritionnelle de 2012 dans le Sahel a été lancée seulement trois mois après la déclaration de la crise de manière à fournir un apprentissage rapide et précoce visant à améliorer la future réponse et à aider à réduire les crises à venir. Cette ETR a permis à l'UNICEF de mettre en place un cadre plus cohérent pour guider sa réponse, notamment des recommandations concrètes sur la manière d'atteindre plus d'enfants en augmentant le nombre de centres fournissant un traitement contre la malnutrition, en utilisant des cliniques mobiles, en assurant la promotion de la gestion des cas de malnutrition aiguë au niveau communautaire, en augmentant les capacités dans les zones isolées et en intégrant mieux la fourniture de services. Les résultats du rapport ont été largement utilisés dans la région : les bureaux pays ont développé des réponses de gestion détaillées immédiatement après l'exercice, et le bureau régional a mis en œuvre les recommandations (UNICEF, 2013).

¹⁸Jeff Crisp, UNHCR, ALNAP Evaluation CoP contribution, juillet 2012.



Regard d'évaluateur 17

L'expérience du PAM en matière d'intégration de l'évaluation dans le développement de politiques

En 2011, le PAM a réalisé une évaluation de sa politique en matière d'alimentation scolaire qui a été conçue pour satisfaire les processus de prise de décision. La nouvelle politique d'alimentation scolaire a été adoptée en 2009 et une condition de son approbation par la direction était qu'elle devrait être évaluée fin 2011, ce qui a permis une planification à l'avance. Cinq études distinctes d'impact pays ont été réalisées et ont nourri l'évaluation de la politique, permettant ainsi qu'elle s'appuie sur de solides fondations avec un calendrier de travail clair. La demande de l'évaluation était également prédéterminée depuis le début afin de s'inscrire dans l'évaluation de la politique par la direction. Le PAM dispose désormais d'une politique prévoyant que les nouvelles stratégies doivent être évaluées après quatre à six ans de mise en œuvre, si bien que la demande future de produits d'évaluation à propos de la stratégie en question est garantie. La même approche a été retenue pour l'évaluation du plan stratégique du PAM. En 2010, le bureau Évaluation du PAM a planifié quatre évaluations stratégiques qui étudiaient toutes différentes dimensions du passage de l'aide alimentaire à l'assistance alimentaire (l'élément central du plan stratégique 2008-2013 du PAM). Cela a nécessité un niveau considérable de pré-planification mais l'utilisation des produits finaux a été considérable.

Source : Sally Burrows, PAM, communication personnelle, mars 2013.

CARE a estimé que l'évaluation/revue après action de la tempête tropicale Jeanne en Haïti était un modèle quant à la manière dont une évaluation peut alimenter de façon efficace une planification et une préparation en partie grâce à l'attention accordée à son timing. Une revue initiale six mois après la tempête a permis la participation d'un bon échantillon représentatif du personnel bien que certains aient déjà quitté le terrain. Dans le même temps, la distribution du rapport final d'évaluation en mars a permis son utilisation lors de l'événement annuel de planification de CARE Haïti en avril de la même année. Ce rapport identifiait les ressources manquantes comme les points de stockage et de distribution d'eau potable, éléments que la séance de planification a permis de résoudre pour la prochaine année fiscale. Enfin, le rapport a été utile pour construire des scénarios et les plans de contingence qui les accompagnent (Oliver, 2007 : 14).

Développer une approche stratégique pour sélectionner ce qui devrait être évalué

Chaque organisation dispose de compétences limitées pour commander, mettre en œuvre des évaluations et en tirer des enseignements. Certaines organisations commanditent des évaluations pour un grand nombre de leurs programmes mais se retrouvent ensuite en difficulté pour assurer la qualité du processus. Même lorsque la qualité est assurée, une réflexion réelle sur les résultats d'un rapport d'évaluation prend du temps. En outre, le fait de prendre en compte et mettre en œuvre les recommandations est encore plus exigeant (encadré 7).

Encadré

7

Une évaluation est-elle toujours justifiée? Quelques éclairages de Carol Weiss

Si les résultats d'une évaluation n'ont pas d'influence sur la manière dont les personnes voient le programme, c'est sans aucun doute un exercice vain. Dans l'un de ses ouvrages de référence, Carol Weiss analyse quatre types de circonstances dans lesquelles une évaluation peut ne pas être un exercice utile.

1. Lorsque le programme connaît une certaine routine et peu de stabilité : si le programme et la manière dont il se déroule ne sont pas clairement définis, l'évaluation n'aura pas non plus un sens clair.
2. Lorsque les personnes impliquées dans le programme n'arrivent pas à se mettre d'accord sur ce que le programme cherche à atteindre.
3. Lorsque le financeur ou le responsable de l'évaluation fixe des limites strictes sur ce que l'évaluation peut étudier, excluant certaines questions importantes.
4. Lorsque qu'il n'y a pas suffisamment d'argent ou de personnel qualifié pour réaliser l'évaluation.

De façon plus positive, Weiss liste également ce qu'elle appelle les « objectifs d'évaluation pour la décision » qui incluent les éléments suivants :

- Corrections à mi-parcours.
- La poursuite, l'extension, l'institutionnalisation, la réduction, la fin ou l'abandon d'un programme.
- Tester une nouvelle idée de programme.
- Choisir la meilleure des alternatives.
- Décider s'il faut poursuivre le financement.
- Comprendre la logique d'intervention du programme.
- Redevabilité pour le financement reçu et les ressources utilisées.

Source : Weiss, 1998: 24-28.

Certains départements d'évaluation examinent les dépenses annuelles de leur organisation et planifient des évaluations de manière à s'assurer que tous les principaux programmes sont régulièrement évalués. De telles approches ont pour fil conducteur le fait que la redevabilité demande que les activités de l'organisation soient couvertes de façon cyclique. Cependant, une approche aussi mécanique ne permet pas nécessairement de garantir une utilisation efficace et un impact aux EHA. De même, le fait d'allouer un pourcentage fixe des coûts d'un programme à des budgets d'évaluation peut ne pas être le moyen le plus rentable d'utiliser des ressources rares.

Il peut être préférable de choisir une approche stratégique (Encadré 8) accordant proportionnellement plus de ressources à des programmes complexes ou plus problématiques et où les preuves d'efficacité sont moins facilement accessibles. Cela peut signifier plus de ressources pour des évaluations concernant des petits programmes plutôt que des grands programmes qui ont duré plusieurs années. L'expérience montre l'intérêt d'essayer d'être plus explicite et intentionnel dans sa sélection des utilisateurs prévus, des utilisations et du timing de l'évaluation. Par conséquent, les organisations peuvent aussi prendre en compte le fait d'utiliser une approche plus intentionnelle pour sélectionner les évaluations et prendre en compte de façon plus explicite les utilisateurs et utilisations prévus de manière à influencer les choix.

Le rapport annuel d'évaluation du PAM de 2011 a observé un intérêt accru pour les évaluations de plus haut niveau : évaluations globales (stratégiques et thématiques), évaluations portant sur un portefeuille pays et série d'évaluations d'impact. Le rapport constatait que, étant donné les ressources restreintes de l'évaluation, cette approche promet une plus grande valeur ajoutée au PAM en fournissant des éléments de preuve alimentant des décisions de niveau stratégique en lien avec les politiques, les stratégies pays ou les stratégies globales alors que les évaluations d'impact fournissent une analyse approfondie des résultats, de l'impact et des effets inattendus plutôt que de simples évaluations portant sur une seule opération.

Encadré

8

Améliorer le ciblage des évaluations au sein de l'UNICEF

En 2012, l'UNICEF a été l'organisation dotée du plus grand nombre de domaines de responsabilités au niveau des Clusters. Elle a ainsi entrepris une « étude systématique, détaillée et indépendante de ses performances dans le cadre de son rôle d'organisation chef de file de Cluster (CLARE) dans l'action humanitaire. La première de ce type fut l'évaluation d'une réponse à une grande variété de pressions internes et externes. Elle visait à produire des preuves de résultats – ou leur absence – en matière de gestion de l'aide pour améliorer la fonction de leadership et coordination du Cluster de l'UNICEF (UNICEF EO, 2012c : 2–4) ».

Un élément important de cette évaluation fut une série d'études de cas pays. De manière à réduire la marge due aux biais et à la subjectivité dans la sélection des pays, l'équipe de l'évaluation humanitaire de l'UNICEF a utilisé une « analyse Cluster » : une technique statistique pour sélectionner les études de cas pays de façon plus systématique et objective. Les bureaux pays ont été regroupés en fonction d'une liste détaillée de variables telles que : le profil de l'urgence (type et niveau), les phases de mise en œuvre du système Cluster (c'est-à-dire activation, mise en œuvre, retrait), le nombre de Clusters en place, le nombre de membres de Clusters aux niveaux national et sous-national, la présence du bureau OCHA, la présence d'une mission de maintien de la paix des Nations unies, les capacités et l'implication de l'autorité nationale de gestion des catastrophes et la disponibilité des fonds d'urgences.

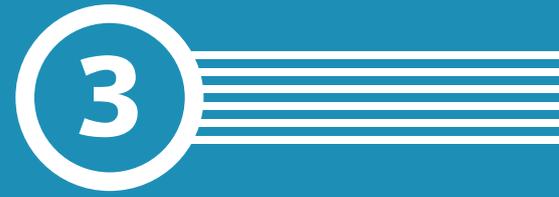
Le bureau Évaluation a alors identifié huit Clusters de bureaux pays avec des caractéristiques et a procédé à la sélection d'un pays par étude de cas pour chacun d'eux (UNICEF EO, 2012b : 1, 3). De plus, le bureau Évaluation s'est efforcé d'assurer la représentation régionale maximale de l'UNICEF et d'éviter une « fatigue de l'évaluation » au sein des bureaux pays les souvent évalués. Cette approche systématique a permis d'obtenir une série d'études de cas pays qui soient plus pertinentes sur le plan thématique et plus représentatives.

Source : Erica Mattellone, UNICEF, communication personnelle, mars 2013.

La revue des pairs du DAC portant sur la fonction Évaluation de DFID a conclu que ses exigences en matière de suivi-évaluation étaient trop complexes, que certaines évaluations étaient trop chronophages et devenaient rapidement obsolètes lorsque les circonstances extérieures changeaient (Riddell, 2009). Cette revue a également observé que le volume des évaluations individuelles de DFID risquait de dépasser les capacités d'absorption des leçons de la plupart des parties prenantes. Un rapport de la Banque mondiale est arrivé à des conclusions similaires :

L'utilité et l'influence de nombreuses évaluations pourtant solides du point de vue méthodologique ont été limitées parce qu'elles étaient vues comme des évaluations au coup par coup et ne faisaient pas partie d'une stratégie systématique de sélection des évaluations répondant aux questions prioritaires en matière de politiques ou en lien avec le budget national et la planification stratégique. (Banque mondiale, 2009 : 70).

DOMAINE DE CAPACITÉS



LES PROCESSUS ET SYSTÈMES D'ÉVALUATION

Renforcer les processus préévaluation	62
Améliorer la qualité des évaluations	64
<i> limiter les objectifs de l'évaluation</i>	64
<i> Impliquer les bénéficiaires</i>	65
<i> L'assurance qualité</i>	66
<i> S'engager dans une revue entre pairs de la fonction Évaluation</i>	68
Diffuser les résultats de façon efficace	70
Renforcer les processus de suivi et post-évaluation, y compris les relations entre l'évaluation et la gestion plus générale des connaissances	75
<i> Garantir qu'il existe une réponse de la direction aux évaluations</i>	78
<i> Garantir un accès et une facilité de recherche vis-à-vis de l'information évaluative</i>	80
Réaliser des méta-évaluations, des synthèses d'évaluation, des méta-analyses et des bilans de recommandations	82

Renforcer les processus préévaluation

Renforcer les processus préévaluation est vital. Une publication de la Banque mondiale qui appelle cela « analyse préalable » la décrit de la façon suivante :

De nombreux évaluateurs sont impatients de finaliser la planification de l'évaluation et, par conséquent, de se lancer dans la collecte des données. Ils essaient de commencer le travail exploratoire en même temps que la collecte des données. Toutefois, finaliser une bonne analyse préalable est essentiel pour comprendre une intervention. Cela peut faire gagner du temps et de l'argent à l'évaluation, permettre que l'évaluation satisfasse les besoins du client, que des relations se maintiennent ou se construisent, pas seulement avec le client mais aussi avec des parties prenantes clés. Plus important, une bonne analyse préalable peut permettre que l'évaluation réponde aux bonnes questions pour obtenir l'information nécessaire plutôt qu'elle ne collecte des données qui ne seront peut-être jamais utilisées.

(Morras Imas et Rist, 2009: 142)

En plus de noter l'importance de l'implication de l'utilisateur et du timing, l'étude de la Banque mondiale recommande de prendre en compte les questions suivantes (Morras Imas et Rist, 2009) :

- En combien de temps l'évaluation doit-elle être réalisée ?
- De quelles ressources dispose l'évaluation ?
- Les apports théoriques des sciences sociales sont-ils pertinents pour cette évaluation ?
- Qu'ont trouvé des évaluations de programmes similaires ? Quelles problématiques ont-elles fait ressortir ?
- Quelle est la théorie du changement derrière le projet, le programme ou la politique ?
- Quelles données existantes peuvent être utilisées pour cette évaluation ?

L'une des conditions les plus importantes en matière de préévaluation est que le programme en train d'être évalué soit évaluable. Cela demande que les responsables du programme s'assurent qu'un cadre de performance est en place avant qu'un programme ne soit mis en œuvre. Un tel cadre devrait, selon les mots de Riddell (2009), clarifier quel

type de réussite est attendu, identifier et spécifier les éléments de preuve quantitatifs et/ou qualitatifs qui seront suivis et utilisés pour juger la performance, y compris l'optimisation des ressources, et répondre à la question de l'attribution. Il en va de même pour le développement des politiques et des stratégies.

En observant par exemple DFID, Riddell a remarqué que dans le passé, l'une des faiblesses de l'approche de DFID était le manque de prise en compte de la future intégration de revues ou évaluations lorsque des politiques et des stratégies clés étaient élaborées et (finalement) lancées. Toutes tendaient donc à être ex-post par nature. Cela a changé à présent, DFID adoptant une approche prospective de l'évaluation de sorte que toutes les nouvelles interventions impliquent une décision pour savoir si une évaluation indépendante est nécessaire dès le commencement du projet. Il est très probable que toutes les interventions de plus de 5 millions de livres, ou encore celles qui sont inédites ou innovantes, intégreront dès le début une évaluation conséquente qui sera également prise en compte dans le projet financier. Le but de ces changements était d'assurer à la fois l'assurance qualité et l'« évaluabilité » des programmes financés par DFID avant leur mise en œuvre (Government Office for Science, 2011a : 14 ; York, 2012 : 19).

En 2012, l'UNICEF a reçu un financement de DFID pour dynamiser ses capacités en matière d'action humanitaire. De même qu'il était attendu que l'UNICEF obtienne des résultats dans quatre domaines en lien avec des compétences spécifiques, il lui était également demandé d'être capable de le démontrer (UNICEF, 2013 : 12–13). Le bureau Évaluation a commandité une étude d'« évaluabilité » posant les questions suivantes :

- La logique du programme était-elle claire ?
- A-t-elle été clairement communiquée aux parties prenantes ? Comprise et acceptée par celles-ci ?
- Les activités ont-elles été planifiées de façon réellement stratégique ? Les ressources étaient-elles suffisantes ?
- Les modalités de gestion étaient-elles optimales ?
- Les données collectées peuvent-elles démontrer que les résultats sont en train d'être atteints ?



Cette étude a aidé la direction à renforcer les liens entre les activités du programme et les résultats ciblés. Elle a également permis un processus plus stratégique d'allocation des fonds du programme. Cet exercice a démontré à l'UNICEF comment l'évaluation et la gestion peuvent fonctionner ensemble pour répondre à des questions avant qu'elles ne deviennent des problèmes (ibid.).

Améliorer la qualité des évaluations

Délivrer des produits d'évaluation de grande qualité peut stimuler un « appétit » de l'organisation pour maintenir des standards de qualité très élevés en matière de travail évaluatif.

La qualité est essentielle pour assurer crédibilité et utilité des évaluations. Elle est manifeste dans l'utilisation précise et adaptée des critères d'évaluation, la présentation des éléments de preuve et l'analyse professionnelle, la cohérence des conclusions avec les résultats de l'évaluation et le niveau de réalisme des recommandations.

(PAM, 2008 : 9)

Le nouveau guide pilote d'ALNAP Évaluer l'action humanitaire approfondit plus largement ces questions (Buchanan-Smith et Cosgrave, 2013). Ce chapitre ne cherche pas à répéter ce travail mais souligne certains thèmes structurants qui ont émergé des entretiens et de la recherche.

Limiter les objectifs de l'évaluation

- Le CICR demandait par le passé à toutes les parties intéressées de lister les questions et sujets qu'elles aimeraient voir abordées dans une évaluation. Le département Évaluation les reformulait ensuite en questions « évaluables ». Toutefois, le périmètre des évaluations s'élargissait sans cesse jusqu'à ce que le processus devienne difficile à gérer. Pour réduire ce problème, le département Évaluation essaie à présent de se focaliser sur seulement trois questions pour chaque évaluation.
- Une étude d'ICAI portant sur la qualité des évaluations de DFID a remarqué que certaines évaluations n'étaient pas assez focalisées et essayaient de couvrir trop de sujets sans priorisation. Cela a eu pour résultat des rapports d'évaluation très longs, ce qui limitait leur capacité à être utilisés (Drew, 2011).

Impliquer les bénéficiaires

Comme évoqué dans le chapitre 2 ci-dessus, le fait d'impliquer les parties prenantes et plus spécifiquement les utilisateurs des évaluations améliore l'utilisation. Toutefois, les personnes affectées par une crise doivent être au centre non seulement du processus d'évaluation mais aussi de l'action humanitaire (IASC, 2012). Bien qu'il reste encore beaucoup de chemin à faire pour atteindre cet objectif, les organisations humanitaires travaillent pour améliorer la situation actuelle (Taylor *et al.*, 2012). Il devient petit à petit la norme que les populations affectées soient impliquées à une étape du processus d'évaluation même si c'est en général seulement comme informateurs plutôt que pour déterminer le programme de l'évaluation. Néanmoins, cela représente une amélioration de la situation par rapport à dix ans en arrière lorsque peu d'évaluations humanitaires comportaient des discussions structurées avec la population affectée¹⁹. Au-delà du fait d'impliquer les populations affectées, certains exercices d'évaluation ont souligné l'importance d'impliquer les partenaires du gouvernement hôte (Encadré 9).

Encadré

9

Tirer le maximum des ateliers d'évaluation inter-acteurs sur le terrain

Dans le cadre de l'évaluation en temps réel (ETR) du Comité permanent inter-agences (IASC) qui portait sur la réponse aux inondations au Pakistan (IASC, 2011a), quatre ateliers provinciaux ont eu lieu à Karachi, Punjab et Islamabad (pour les provinces KPK et Baluchistan). Entre 20 et 40 personnes provenant du gouvernement, des ONG internationales et locales, des organisations des Nations unies et du mouvement Croix-Rouge ont participé à chaque atelier. Il y eut ensuite un atelier national impliquant des bailleurs et certains représentants d'organisations qui avaient participé aux ateliers provinciaux. Des recommandations étaient reprises d'un atelier à un autre et étaient améliorées à chaque occasion. Cela a eu pour effet de construire un consensus, d'affiner les recommandations et la clarté des priorités, le timing et de clarifier les rôles de chacun et à quel niveau (provincial, national ou global).



¹⁹Voir par exemple :
Catley *et al.*, 2008 ;
ALNAP et Groupe
URD, 2009 ;
Pankaj, Welsh *et*
Ostenso, 2011 ;
Anderson, Brown *et*
Jean, 2012.



Cette ETR du IASC était la première à utiliser des ateliers comme principal mécanisme de restitution, ce qui a permis aux personnes sur le terrain de participer activement à l'élaboration des recommandations tout en faisant en sorte que ce processus ne soit plus un exercice imposé par le siège (www.youtube.com/watch?v=qW-XR14k4NQ). Il semble important que la plupart des recommandations de l'ETR IA du Pakistan aient par la suite été adoptées. Les participants ont trouvé ce processus utile pour réfléchir attentivement à ce qu'ils devaient faire mais aussi pour dynamiser l'apprentissage en temps réel et la redevabilité entre pairs sur le terrain. L'implication du gouvernement dans ce processus a également contribué à établir des priorités pour différentes actions de suivi.

Chaque atelier a duré plus d'une demi-journée. Le processus total des ateliers a pris une semaine entière du temps de l'évaluation mais il a été considéré comme très utile et un élément clé du consensus obtenu et de la mise en œuvre des recommandations. De plus, les recommandations pouvaient être mises en œuvre immédiatement sans avoir à attendre que le rapport d'évaluation final ne soit publié. Une fois approuvées et priorisées sur le terrain, les conclusions et les recommandations ont été présentées à Genève et à New York, ce qui a permis une diffusion rapide et une appropriation politique au niveau du siège. En complément avec d'autres évaluations inter-agences clés, cette ETR a alimenté l'Agenda Transformatif du IASC (IASC, 2012).

Source : Riccardo Polastro, DARA, contribution à la CdP Évaluation d'ALNAP, juillet 2012.

L'assurance qualité

Un certain nombre d'organisations cherchent à améliorer la qualité de leurs évaluations et à standardiser les approches au moyen de systèmes d'assurance qualité et/ou de processus de revues externes. Le PAM utilise une approche appelée Système d'assurance qualité de l'évaluation (EQAS en anglais) qui se concentre sur les résultats immédiats de l'évaluation et donne aux équipes d'évaluation des conseils clairs sur les standards attendus. Cela donne également au personnel en charge des évaluations des conseils précis en matière de timing et de standards à appliquer. EQAS ne s'intéresse pas actuellement à de plus larges processus autour de l'évaluation mais une revue est en cours et celle-ci étudiera comment élargir le processus d'assurance qualité aux systèmes de gestion des évaluations comme la sélection des équipes, la réalisation des processus post-évaluation et le fait de s'assurer que les opportunités d'apprentissage sont optimisées.

Bien que les responsables d'évaluation aient pris du temps pour l'adopter totalement, EQAS est en train d'être reconnu comme un atout pour l'organisation. Les équipes d'évaluation semblent apprécier la clarté des attentes qu'il apporte dès le début d'une évaluation bien que certains se soient battus contre ce qu'ils ressentaient comme un manque de flexibilité. D'autres organisations des Nations unies ont également exprimé un intérêt pour adopter une approche similaire (PAM, 2009)^[20].

Les exigences standardisées d'EQAS ont amélioré la qualité des rapports d'évaluation et le processus collaboratif utilisé pour développer les supports a augmenté la compréhension et l'application de ces standards. En 2011, le bureau Évaluation a étendu son usage des relecteurs externes pour ces évaluations avec des niveaux d'intérêt particulièrement élevés pour les différentes parties prenantes. Ces comités de relecture sont distincts des consultants indépendants qui conduisent les évaluations et apportent une dimension supplémentaire à l'assurance qualité de la méthodologie et/ou du contenu. (PAM, 2011)

²⁰Sally Burrows,
WFP, communication
personnelle, mars
2013.



Regard d'évaluateur 18

L'assurance qualité et le système de supervision des évaluations de l'UNICEF

L'UNICEF a introduit un système de supervision générale des rapports d'évaluation^[21] (GEROS en anglais) qui fournit à la direction générale un examen clair et synthétique de la qualité et de l'utilité de chaque rapport d'évaluation. Ce système cherche également à renforcer les compétences en évaluation en apportant des retours en temps réel sur la façon d'améliorer les évaluations futures. Enfin, il contribue à la gestion des connaissances et à l'apprentissage de l'organisation en identifiant les rapports d'évaluation de qualité satisfaisante à utiliser dans le cadre d'une méta-analyse. L'examen des rapports finaux d'évaluation est externalisé auprès d'une société indépendante. Un rapport d'évaluation est jugé comme satisfaisant lorsqu'il se fonde sur des éléments de preuve crédibles et répond aux objectifs de l'évaluation, pouvant dès lors être utilisé en toute confiance.

L'un des critères utilisés pour juger du caractère satisfaisant d'un rapport d'évaluation est que l'objectif de l'évaluation soit clairement défini, notamment la raison pour laquelle l'évaluation a été nécessaire à un certain moment, qui a eu besoin des informations, quelles étaient les informations demandées et comment ces informations allaient être utilisées. Le personnel de l'UNICEF en charge de l'évaluation a observé que, lorsque GEROS jugeait qu'une évaluation était de haute qualité, les responsables avaient davantage confiance pour agir sur les résultats, mais aussi que leur engagement vis-à-vis des évaluations et de la mise en œuvre des recommandations était renforcé.

Source : Rob McCouch, UNICEF, communication personnelle, mars 2012.

²¹UNICEF, 2010c; see also: www.unicef.org/evaluation/index_60830.html

S'engager dans une revue entre pairs de la fonction Évaluation

Ces dernières années, et particulièrement au sein des agences des Nations unies, le nombre de revues entre pairs de la fonction Évaluation menées a augmenté, revue au cours de laquelle des collègues d'organisations similaires viennent évaluer le dispositif et la performance générale du bureau Évaluation. Cependant, au niveau des ONG, ces revues entre pairs sont encore rares pour évaluer des organisations partenaires ou sœurs. Cela se produit pourtant dans d'autres domaines d'intérêt des ONG.^[22] La revue entre pairs des

²²Par exemple : la performance (NGO and Humanitarian Reform Project, 2010), ainsi que les questions programmatiques (IFRC, 2011 ; Cosgrave et al., 2012 ; CBHA, 2012) et thématiques comme la redevabilité (SCHR, 2010).

Nations unies s'engageait à étudier trois sujets centraux : la crédibilité, l'indépendance et l'utilité (UNEG, 2005), comme présenté dans l'Encadré 10. L'objectif de ces revues entre pairs a été de créer plus de connaissances, de confiance et d'utilisation des systèmes d'évaluation par les cadres, les organes directeurs et d'autres personnes, de proposer une manière appropriée d'« évaluer les évaluateurs » et de partager les bonnes pratiques, les expériences et l'apprentissage mutuel (DAC/UNEG, 2007 ; UNEG, 2011).

Encadré**10****Les trois grands principes de l'évaluation**

En 2008, le cadre de l'UNEG pour la revue entre pairs a proposé trois principes centraux en ce qui concerne l'évaluation : l'indépendance, la crédibilité et l'utilité.

La crédibilité a trait à la compétence des évaluateurs, à la transparence du processus d'évaluation et à l'impartialité des évaluateurs et du processus. L'indépendance signifie que l'évaluation est libre de pressions politiques ou organisationnelles, ou encore de préférences personnelles qui pourraient biaiser sa conduite, ses résultats, ses conclusions ou ses recommandations. Elle implique que les évaluations sont réalisées ou gérées par des entités ou des individus non contrôlés par des personnes responsables de la conception et de la mise en œuvre du sujet de l'évaluation.

L'utilité signifie que les évaluations visent à influencer la prise de décision et y parviennent. Les évaluations doivent être pertinentes et utiles mais aussi présentées de façon claire et concise. Elles ont une valeur dans la mesure où elles servent les besoins d'informations et de prises de décision des utilisateurs prévus, notamment en répondant aux questions soulevées à propos de l'évaluation par les personnes qui l'ont commanditée.

Source : Heider, 2011 (d'après UNEG, 2005 ; UNEG, 2008 ; PAM, 2008).

La première revue entre pairs des Nations unies s'est déroulée en 2005 et cette pratique est depuis un élément régulier du travail de l'UNEG. Le processus a été initialement conçu pour traiter le fait que les bailleurs bilatéraux ne donnaient pas l'impression d'utiliser les rapports et les résultats des travaux d'évaluation réalisés par les agences multilatérales (UNEG, 2013). L'objectif était donc d'asseoir la crédibilité des rapports d'évaluation provenant de l'organisation elle-même et, par conséquent, de réduire le besoin d'évaluations externes multi-bailleurs d'une agence ou de son bureau Évaluation. Le principal avantage a été l'apprentissage et l'échange entre pairs et, en particulier, l'opportunité pour les évaluateurs de différentes agences d'apprendre des réalités et des pratiques des autres. Améliorer le statut et souligner la valeur de la fonction Évaluation dans une organisation donnée, particulièrement aux yeux des équipes Stratégie et Politiques mais aussi des organes de direction, ont également été vus comme des bénéfices importants (ibid : 13).

Diffuser les résultats de façon efficace

Les produits d'évaluation, et en général les informations issues des évaluations, intéressent potentiellement des publics très divers pouvant comprendre : les participants des programmes et les communautés affectées, les responsables des programmes, les bailleurs, les conseillers, les décideurs, les représentants gouvernementaux, les chercheurs et les universitaires, les médias locaux et internationaux, etc^[23]. Idéalement, chaque public doit recevoir un support de communication qui présente et synthétise les informations évaluatives en fonction de ses besoins spécifiques en matière d'apprentissage et d'information (Hallam, 1998 ; Proudlock et Ramalingam, 2009 : 13 ; Morra Imas et Rist, 2009 : chapitre 13 ; Buchanan-Smith et Cosgrave, 2013 : chapitre 7).

En ce sens, les organisations auront peut-être envie de mettre en place des politiques claires en ce qui concerne la personnalisation des produits informatifs ainsi que la communication et la diffusion des documents issus de l'évaluation. Par exemple, comme exposé dans le « Cadre d'évaluation » de la FICR, les TdR doivent comprendre une liste de diffusion initiale « pour assurer que le rapport ou le résumé de l'évaluation atteint son public prévu... La diffusion du rapport d'évaluation peut prendre différentes formes appropriées en fonction du public spécifique » (FICR, 2011a : 15). Cela implique une version interne et une version externe du rapport. La diffusion stratégique des résultats est essentielle pour rendre les évaluations plus efficaces.

Dans le cadre d'une communication personnelle, un évaluateur de rang élevé a commenté la différence d'impact des évaluations qu'il avait remarquée en passant d'un grand département Évaluation disposant de solides ressources à un département bien plus petit au sein d'une autre organisation. Malgré la différence au niveau des ressources disponibles, les évaluations réalisées par la plus petite organisation ont un impact bien plus significatif. Plusieurs raisons expliquent cela, la plus importante étant l'engagement de la plus petite organisation pour diffuser les résultats et orienter cette diffusion de manière efficace. Le processus d'évaluation commençait en effet avec la stratégie de communication, la diffusion étant réfléchie avant même que le rapport (peut-être inapproprié) ne soit disponible. Le fait de planifier la stratégie de diffusion du rapport en avance influence le type d'informations collectées tout au long de l'évaluation et la manière dont elles sont présentées, ce qui contribue à garantir qu'elles répondent aux besoins d'informations des décideurs.

Jones et Mendizabal (2010) partagent les résultats présentés ci-dessus puisqu'ils ont remarqué que la manière et le moment de présenter les éléments de preuve ont une réelle importance. Parmi les solutions, on peut retenir : des séminaires et des présentations ciblés, des briefings individuels pour les chefs d'équipe, une newsletter ou des documents préparatoires du département Évaluation, des formats adaptés aux e-mails et le développement de nouveaux supports comme les documents qui présentent les leçons tirées des évaluations selon des axes thématiques, régionaux/nationaux et programmatiques/politiques. Il a également été jugé important de se concentrer sur les interrelations personnelles plutôt que via des intermédiaires. Enfin, la même étude a observé que le format et la présentation des évaluations pouvaient être améliorés : « Certaines personnes interrogées avaient le sentiment que les rapports complets étaient trop longs et trop techniques... Il est probable qu'il s'agisse d'un problème général au sein du secteur de l'évaluation ». (voir Encadré 11)



La diffusion stratégique des résultats est essentielle pour rendre les évaluations plus efficaces.

Encadré

11

Sept nouvelles manières de communiquer les résultats d'une évaluation

1. Fiches de synthèse ou notes de recherche et de politique. Un document plus court est bien plus susceptible d'être lu qu'un rapport complet.
2. Tableaux de résultats. Il existe un risque de simplification mais le fait de présenter les résultats de façon brute peut transmettre votre message de façon très forte.
3. Fiches de score ou tableaux de bord pour un suivi en temps réel.
4. Rapports interactifs, pages Internet interactives ou applications Internet (par exemple : <http://www.usahidi.com/>).
5. Romans-photos, bandes dessinées, graphiques d'information, illustrations.
6. Des blogs peuvent être utilisés durant le processus d'évaluation, notamment pour débattre de son utilisation.
7. Reportage video/multimédia.

Source : d'après O'Neil, 2012

**Regard d'évaluateur 19****Le grand impact des rapports de fin de mission**

Un évaluateur remarque que la partie la plus utile du processus d'évaluation est le rapport de fin de mission. Bref, passionné et dépourvu du « remplissage » que l'on retrouve dans les rapports de fin d'évaluation, il est lui aussi disponible pour tous ceux qui sont en train d'être évalués et ce, très rapidement après l'événement. Grâce à cela, et parce qu'il n'est pas rendu public, il est souvent susceptible de générer un véritable débat et une réponse au sein des équipes pays impliquées. A contrario, la circulation du rapport final ne produit jamais autant de débat et d'apprentissage.

Source : Luz Gómez Saavedra, Oxfam Intermón, communication personnelle, mars 2013.

Les entretiens réalisés dans le cadre de cette étude ont révélé certaines stratégies de diffusion innovantes.

- Plusieurs organisations, dont l'UNHCR, le CICR et le Groupe URD, ont réalisé des documentaires portant sur des évaluations, certains rendant compte avec force des questions clés soulevées par l'évaluation uniquement grâce aux paroles des bénéficiaires.

- USAID a commandité un film sur le processus d'évaluation (Encadré 11).
- Le réseau ALNAP a lui-même réalisé une série de documents thématiques de capitalisation suite à l'analyse d'évaluations. Ceux-ci sont bien reçus : les documents portant sur la réponse aux tremblements de terre ont été téléchargés plus de 3500 fois dans les jours qui ont suivi le séisme de janvier 2010 en Haïti.
- Une façon ludique de présenter des résultats consiste à utiliser l'approche Ignite dans laquelle toutes les présentations durent 5 minutes maximum, chaque nouvelle diapositive apparaissant automatiquement à l'écran toutes les 15 secondes. C'est incroyable tout ce qu'on peut dire en cinq minutes !
- Certains évaluateurs ont désormais des blogs où ils présentent leur travail ; de même, Facebook et Twitter suscitent de plus en plus d'intérêt pour véhiculer des messages d'évaluation. (Toutefois, une bande passante faible et un accès Internet sont des problèmes critiques dans certains bureaux pays. Même là où cela n'en est pas un, la surcharge d'information peut aussi être un handicap.)

Encadré

12

Les expérimentations d'USAID en matière de films et les apprentissages réalisés à partir du travail de terrain des évaluations

Le bureau Apprentissage, Évaluation et Recherche d'USAID a récemment engagé un vidéaste, Paul Barese, chargé d'accompagner une équipe d'évaluation sur le terrain. Le réalisateur a suivi une équipe d'évaluation de projet durant plus de trois semaines afin de documenter la plupart des phases de l'évaluation. Un des objectifs était de générer de l'apprentissage à propos du processus d'évaluation, de faire ressortir certains défis rencontrés sur le terrain (en particulier dans les zones affectées par un conflit) et de permettre à ceux qui gèrent, commanditent, conçoivent et réalisent des évaluations de mieux comprendre la complexité et la valeur d'un tel exercice.

La vidéo est disponible à l'adresse suivante : www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program.

Source : Paul Barese in AEA 365 <http://aea365.org/blog/?p=8768> Linda Morra Imas, IDEAS Evaluation listserv exchange, avril 2013.

Encadré

13

Choisir des stratégies de diffusion de l'évaluation

Les facteurs à prendre en compte au moment de mettre en place des stratégies de diffusion sont :

- **Publics** – Quels sont les publics qui recevront le document ? Quels sont leurs besoins particuliers ?
- **Justification** – Pourquoi faites-vous cela ? Que souhaitez-vous faire ?
- **Contenu** – Quel type d'informations contiendra-t-il ?
- **Objectif/utilisation** – Pourquoi est-ce nécessaire ? Pour quoi pourrait-il être utilisé ?
- **Calendrier** – Quand sera-t-il terminé ? Quand sera-t-il distribué ?
- **Développement** – Que reste-t-il à faire pour que le document soit finalisé ?
- **Questions/contraintes particulières** – Pourquoi ce document est-il important – doit-il être accompagné par d'autres informations ? Certains éléments sont-ils nécessaires pour accéder au document ou le comprendre ?
- **Distribution** – De quelles façons sera-t-il distribué ? Différentes personnes le recevront-elles de façon différente ?
- **Préoccupations particulières** – Quelles sont les limites de ce format ?

Ces facteurs ont été utilisés pour analyser six stratégies de diffusion : rapport d'évaluation complet, résumé exécutif, présentation PowerPoint, fiche d'information, bilan statistique et outil en ligne. Il est apparu que ces différents produits de diffusion ont différents avantages et inconvénients. Le rapport complet, par exemple, a été lu par très peu de personnes alors qu'il a été utilisé comme une ressource pour aider à résoudre des sujets conflictuels. D'autres produits comme les fiches d'information de trois pages ont été très appréciés par les collecteurs de fonds en tant que support permettant de montrer ce qui a été réalisé alors qu'elles n'ont pas du tout été lues par les personnes chargées de la mise en œuvre. Il a été observé que plus le public est divers, plus le format de diffusion doit être individualisé. Les présentations orales se sont révélées efficaces, en particulier lorsqu'elles encourageaient la discussion des résultats.

Source : Lawrenz et al., 2007

Même lorsqu'un rapport d'évaluation classique est conservé, des mesures peuvent être prises pour améliorer ses chances d'être utilisés. Une étude de CARE suggérait que, de même que les rapports d'évaluation devraient être plus brefs et plus précis, une feuille d'accompagnement devrait également être rédigée pour les rapports d'évaluation. Celle-ci

classifierait les leçons apprises par domaines spécifiques (comme les ressources humaines, les relations extérieures ou l'approvisionnement) afin de faciliter l'utilisation des résultats par les personnes en charge de ces domaines (Oliver, 2007 : 18).

La même étude a également observé que les rapports n'étaient pas partagés de façon systématique et que très peu des personnes interrogées semblaient savoir où trouver un stock de rapports. Il apparaît ainsi de façon claire que même si le rapport est magnifiquement présenté, son usage sera limité si les personnes ne peuvent pas le trouver ! Un format standardisé a également été considéré comme utile par ceux qui souhaitent pouvoir parcourir rapidement le rapport (il aiderait également les évaluateurs à s'assurer que leurs produits sont conformes aux attentes de CARE). Les champs de texte du format pourraient être reliés à une base de données dotée d'un moteur de recherche pour permettre d'accéder facilement aux leçons apprises dans un format concis, soit à partir d'une seule évaluation, soit sous forme de synthèse (par exemple, un résumé des recommandations traitant des ressources humaines sur les deux dernières années).

Renforcer les processus de suivi et post-évaluation, y compris les relations entre l'évaluation et la gestion plus générale des connaissances

L'évaluation est seulement l'un des outils avec lequel une organisation peut promouvoir l'apprentissage et le changement ou encore enregistrer de nouvelles pratiques, des connaissances et une mémoire institutionnelle. Une récente publication de l'UNICEF sur sa pratique de l'évaluation humanitaire (UNICEF, 2013) observe qu'une fonction Gestion de la connaissance est nécessaire pour aider le personnel à apprendre et agir à partir des leçons d'évaluations et d'autres sources de connaissances. Le département des opérations d'urgence de l'UNICEF (EMOPS) a essayé d'intégrer l'évaluation dans le cadre des initiatives politiques de renforcement de l'apprentissage et de la redevabilité en situation d'urgence, en incluant par exemple la dimension de l'évaluation dans des travaux de groupes sur la redevabilité et la gestion des connaissances. L'UNICEF reconnaît que le bureau Évaluation doit avoir les moyens de fournir des conseils stratégiques au-delà de la gestion des évaluations demandées. Cela nécessite également des capacités propres de gestion des connaissances.

Comme le remarque Sandison dans son étude ALNAP sur l'utilisation de l'évaluation, les « planificateurs d'évaluation et les évaluateurs pourraient en faire plus pour reconnaître la relation entre une évaluation et d'autres processus d'apprentissage » (Sandison, 2006 : 135). Sans chercher à couvrir la vaste littérature relative à la gestion de l'apprentissage et des connaissances (voir par exemple Hovland, 2003 ; Ramalingam, 2005), les éléments ci-dessous évoquent certains des sujets qui affectent la force des liens entre évaluation, gestion des connaissances, suivi et utilisation des données évaluatives.

Les recherches et les études de l'UNEG sur le suivi de l'évaluation se sont intéressées à la manière dont la transparence de la réponse de la direction en matière d'évaluation mais aussi des processus de suivi augmentaient le taux de mise en œuvre des recommandations (UNEG, 2010). Le schéma de l'encadré 14 illustre les liens qu'UNEG voit entre les processus d'évaluation et le suivi de l'évaluation comme discuté dans le paragraphe suivant.

Encadré

14

Lier les processus d'évaluation et de suivi

Le schéma ci-dessous permet de visualiser les résultats d'un récent travail de recherche commandité par l'UNEG (2010) dans le but d'identifier quelles sont les préconditions nécessaires pour soutenir une réponse de direction efficace et suivre l'évaluation, à savoir : une planification et une mise en œuvre de l'évaluation de haute qualité. Tout au long de cette étude, un certain nombre d'indications et d'éléments en lien avec des processus d'évaluation de haute qualité ont déjà été présentés. Dans ce contexte, « haute qualité » doit être compris comme un attribut du processus d'évaluation qui favorise et soutient une adoption et une utilisation efficaces de l'évaluation^[25].

²⁵ Une analyse plus complète et plus centrée sur l'évaluation humanitaire concernant ces éléments et les attributs du processus d'évaluation dépasse le cadre de cette étude mais elle peut être trouvée dans la littérature : voir par exemple Rey et Urgoiti (2005) ; ECB (2007) ; Groupe URD (2009) ; IFRC (2011b) ; MSF (2012) ; Cosgrave et Buchanan-Smith (2013).

Préconditions pour un suivi d'évaluation et une réponse de direction efficaces

Bonne planification de l'évaluation

Identification des acteurs clés, définition des objectifs de l'évaluation, préparation des TdR, sélection de l'équipe d'évaluation, organisations logistique pour les missions d'évaluation

Mise en œuvre de l'évaluation de qualité

Briefings et événements de lancement, terrain et travail documentaire, préparation du rapport, processus de commentaires des parties prenantes et contrôle qualité des rapports provisoires (focalisation sur la qualité et la pertinence des résultats, leçons et recommandations)

Réponse de la direction vis-à-vis de l'évaluation

La direction dont les opérations ont été évaluées met en place une réponse, le gouvernement et/ou d'autres partenaires peuvent aussi répondre à l'évaluation

Divulgence et diffusion du rapport d'évaluation

Divulgence et publication (version électronique et/ou papier) de l'évaluation, y compris de la réponse de la direction; résumé de l'évaluation ou autres produits de partage des connaissances/d'apprentissage

Suivi de l'évaluation

Processus formels et informels pour promouvoir et vérifier qu'un apprentissage ancré dans l'évaluation a bien lieu au sein de l'organisation et chez certains partenaires; rapports de la direction sur l'état de la mise en œuvre des recommandations

Garantir qu'il existe une réponse de la direction aux évaluations

Réponse de la direction et suivi des évaluations sont essentiels pour améliorer l'impact de ces dernières (UNEG 2010). Toutefois, il est fréquent dans le langage courant de parler des évaluations « qui prennent la poussière sur les étagères », de rapports qui ne sont jamais lus et d'une quasi-absence de suivi formel ou informel des résultats et des recommandations. Lorsque les mêmes recommandations ne sont pas prises en compte ou sont négligées à chaque fois, que ce soit à cause de leur mauvaise formulation, d'un manque de pragmatisme, de problèmes de calendrier, etc., il est possible que se développe l'impression que les évaluations sont principalement faites à titre symbolique ou pour les apparences^[26].

Il est devenu commun que les organisations aient un système plus formel de réponse de la direction vis-à-vis des évaluations dont le but est de réduire les risques qu'un processus d'évaluation ne se finisse avec le seul rendu du rapport. Différentes organisations ont ici des approches différentes. Le PNUD a créé un Centre de ressources des évaluations, plateforme publique conçue pour améliorer la redevabilité du PNUD et où les plans d'évaluations sont consignés en même temps que la réponse de la direction. Les réponses et le suivi sont observés par le département Évaluation et rapportés au bureau exécutif du PNUD. Certains bureaux pays de DFID organisent une journée de travail spécifique pour passer en revue les cadres de performance et les résultats de l'évaluation de manière à s'assurer que les leçons clés ne sont pas perdues (Hallam, 2011).

À la FAO, il existe un processus au sein duquel l'encadrement supérieur commente la qualité des rapports d'évaluation ainsi que les résultats qu'il accepte et les actions qu'il prévoit pour y répondre. Cette rétroaction est présentée au corps dirigeant en même temps que l'évaluation. De plus, pour les grandes évaluations, le processus comprend une étape supplémentaire : deux ans après l'évaluation, les responsables doivent rendre compte au corps dirigeant des actions entreprises suite aux recommandations acceptées à l'époque. Pour la FAO, cela s'est révélé un puissant outil de gestion qui permet de revisiter les résultats d'évaluation et d'avoir une discussion sur la réponse de la direction (Hallam, 2011).

Depuis 2011, le Secrétariat du PAM organise une consultation informelle sous la forme d'une table-ronde avant chaque réunion du Conseil d'administration, ce qui permet qu'une discussion plus approfondie des rapports d'évaluation et des réponses de la direction soit

.....
²⁶Sur ces points, voir par exemple : Weiss 1998 ; Patton 1997 ; McNulty 2012.

présentée de façon formelle durant cette réunion. Cela a enrichi les interactions entre le Conseil d'administration et la direction sur les sujets soulevés par les rapports d'évaluation et a également permis des échanges plus concis et plus ciblés durant les réunions formelles du Conseil d'administration (PAM, 2011).

La Commission indépendante pour l'impact de l'aide comprend dans son rapport annuel un chapitre qui présente les recommandations de chacun de ses rapports, qu'ils aient été acceptés ou rejetés par DFID, un rapport d'étape sur la mise en œuvre des recommandations et aussi la réponse de DFID suite aux recommandations d'ICAI. Par ailleurs, le DEC demande à ses membres de partager les résumés des évaluations et la réponse de la direction aux évaluations qui seront enregistrés dans la base de données des rapports d'ALNAP (ERD). L'objectif est de créer une culture du partage de l'apprentissage au sein de l'ensemble du secteur humanitaire.



Regard d'évaluateur 20

Les ateliers post-mission du DEC

Pour encourager davantage l'utilisation des enseignements tirés des évaluations, le DEC organise des ateliers post-mission au Royaume-Uni qui peuvent servir à partager des résultats d'une manière générale ou à se focaliser sur des thèmes spécifiques. En 2012 par exemple, le DEC a co-facilité un atelier pour encourager la prise en compte du handicap et de l'âge dans les réponses humanitaires. Il est intéressant de noter qu'il était demandé aux participants de créer un plan d'action personnel qui serait suivi six mois plus tard.

Source : Annie Devonport, DEC, communication personnelle, juin 2013.

Garantir un accès et une facilité de recherche vis-à-vis de l'information évaluative

Passer en revue, filtrer et avoir la possibilité d'accéder de façon efficace à des informations pertinentes en matière d'activités évaluatives actuelles et planifiées constitue certainement une perspective attrayante pour les évaluateurs ; de même pour les informations relatives aux indications, résultats, conclusions et recommandations générées à partir du travail d'évaluation. Le fait de faciliter la collecte et l'accès à de telles informations évaluatives a donc le potentiel pour améliorer les chances que les ressources évaluatives soient dépensées de façon plus efficace et pour réduire la duplication dans le cadre d'évaluations. Cela pourrait également améliorer la probabilité que les résultats et les recommandations des évaluations soient observés, référencés et adoptés de façon plus systématique de manière à alimenter la prise de décision. L'encadré 15 présente trois exemples d'initiatives récentes pour répondre à certains de ces sujets.

Encadré

15

Des exemples d'amélioration de l'accès à l'information en matière d'évaluation

Tearfund a récemment développé un outil Google Docs pour partager avec toute l'organisation les évaluations planifiées et achevées. Avant qu'une évaluation ne débute, le personnel en charge de la gestion ou de la supervision devra fournir des détails sur : les secteurs couverts, le déclencheur de l'évaluation (demande du bailleur, collecte des leçons apprises, etc.), type d'évaluation (moyen-terme, finale ou ETR), composition de l'équipe d'évaluation (interne ou externe), travail de bureau ou de terrain, calendrier approximatif et cohérence avec les résultats de l'organisation. Une fois l'évaluation terminée, des éléments tels que le plan de réponse de la direction, la contribution aux résultats de l'organisation et les leçons clés seront ajoutées. L'attente et le principal objectif de cet outil sont les suivants : a) proposer un instantané de toutes les évaluations planifiées et achevées par Tearfund ; et b) faciliter le partage des connaissances et la communication auprès des personnes impliquées dans des activités d'évaluation et du personnel opérationnel désireux d'avoir accès aux résultats et recommandations des évaluations. L'appropriation de l'outil relève du bureau Suivi-Évaluation au siège de l'organisation^[27].





Le **prototype Humanitarian Genome** lancé récemment cherche à faire passer l'innovation dans la recherche, l'accès et l'extraction des informations relatives aux évaluations. Le projet implique le développement d'un moteur de recherche en ligne qui trie et traite les éléments d'analyse des évaluations humanitaires individuelles. Les rapports d'évaluation ont été numérisés et des passages précis (désignés comme les « éléments d'analyse » ou « insights ») ont été extraits puis codés et notés. La note vise à déterminer l'utilité d'un rapport et son classement dans les résultats de recherche selon un processus similaire à la fonction de classement des recherches de Google ou d'autres moteurs de recherche. Il est à noter que le Humanitarian Genome permet aux utilisateurs de rechercher des informations en lien non seulement avec le contenu de l'évaluation mais aussi avec les processus et méthodes^[28].

En ce qui concerne les bases de données d'évaluations, un des défis consiste non seulement à assurer la facilité de la recherche mais aussi la complétude des supports évaluatifs et des références. Mise en place à la fin des années 1990, la base de données des rapports d'évaluation d'ALNAP (**ERD pour Evaluative Reports Database**) offre à ce jour la somme la plus complète de documents d'évaluations humanitaires : rapports d'évaluation mais également méthodes d'évaluation et documents d'orientation ou encore matériaux choisis sur la recherche dans le domaine de l'évaluation. L'objectif de complétude et de fiabilité des résultats de recherche repose cependant plus sur les membres du réseau qui communiquent leurs rapports d'évaluation. Cela demande également des efforts continus pour rechercher et extraire des données d'évaluations sur d'autres sites de manière à être aussi exhaustif que possible. Le système de codage ou d'étiquetage doit être souvent révisé en pensant aux besoins des utilisateurs de manière à optimiser la facilité de recherche. Cela implique de fournir des options de recherche intuitives et la capacité de filtrer les résultats de recherche^[29].

²⁷Catriona Dejean, Tearfund, communication personnelle, juillet 2013.

²⁸*Results Matter*, 2013. Pour plus de détails, voir : www.humanitarianinnovation.org/projects/large-grants/groningen

²⁹Pour plus de détails, voir : www.alnap.org/resources.

Réaliser des méta-évaluations, des synthèses d'évaluation, des méta-analyses et des bilans de recommandations

La plupart des utilisateurs potentiels des résultats d'évaluations veulent en savoir plus que simplement ce qu'une évaluation ou une étude a trouvé. « Ils veulent connaître le poids de la preuve » (Weiss, 1998 : 317). Des douzaines d'études, d'évaluations et de revues peuvent couvrir le même sujet ou la même thématique. Le fait de les étudier dans leur totalité, via une méta-analyse ou une synthèse des évaluations, peut produire des preuves et des résultats bien plus riches. Par exemple, une étude de la Banque mondiale sur l'utilisation de l'évaluation d'impact pour le développement de politiques a montré que :

Même si des évaluations individuelles peuvent avoir été d'une grande utilité, les cas illustrent que les effets et les bénéfices sont souvent cumulatifs et que l'utilisation et la participation du gouvernement tendent à augmenter là où il existe une série d'évaluations. Dans plusieurs cas, la première évaluation était insuffisante du point de vue méthodologique... mais lorsque les résultats ont été trouvés utiles par les partenaires nationaux, cela a généré de la demande pour d'autres évaluations plus rigoureuses.

(Banque mondiale, 2009 : 70)

La demande de synthèses d'évaluations, de revues thématiques et de méta-analyses dans le secteur humanitaire semble être croissante. Tous les deux ans, le Secrétariat d'ALNAP demande aux auteurs du rapport L'État du système humanitaire de conduire une méta-analyse de tous les rapports d'évaluation humanitaire collectés et stockés, sur cette période de deux ans, dans la base de données des rapports d'évaluation d'ALNAP (voir Encadré 15). La méta-analyse est le point de départ de la recherche et de l'analyse suivantes qui étudient, et suivent au fil du temps, les tendances et les performances de l'ensemble du système humanitaire d'après les critères OCDE-DAC.

D'autres organisations ont exploré des approches consolidées en matière d'analyse des évaluations :

- Oxfam GB a commencé à conduire de façon régulière des méta-évaluations de ses actions en utilisant une série de critères (Oxfam GB, 2012 ; n.d.).
- CARE a réalisé une méta-revue d'évaluations et de revues après-action à partir de 15 réponses d'urgence et en a rassemblé des apprentissages clés (Oliver, 2007 : Annexe 1A).
- Chaque année, NORAD produit un rapport de synthèse des leçons tirées de ses évaluations^[30].
- Une publication d'ODI sur les leçons tirées des recherches et évaluations au sein de DFID (Jones et Mendizabal, 2010) a remarqué que les personnes interrogées ont été très enthousiastes à propos de travaux de recherche et d'évaluation qui synthétisent et comparent différents études sur des thèmes particuliers.
- En 2012, ACF a mis en place un projet d'apprentissage des réponses d'urgence (Encadré 16).
- Le mouvement Croix-Rouge Croissant-Rouge a tenu en 2012 une Conférence d'apprentissage à l'échelle de l'organisation sur son opération en Haïti^[31]. L'édition de 2013 comprendra une méta-analyse de toutes les évaluations de programme et techniques ainsi que des revues réalisées par la FICR, la Croix-Rouge haïtienne ou toute autre société nationale ayant participé aux opérations qui ont suivi le séisme en Haïti. L'objectif est d'utiliser les synthèses d'évaluations pour enregistrer les bonnes pratiques et les améliorations techniques susceptibles d'alimenter de futures mises à jour des procédures opérationnelles permanentes du mouvement Croix-Rouge Croissant-Rouge (FICR, 2013).
- ACIDI a commencé à publier un rapport annuel sur les leçons tirées des évaluations durant l'année écoulée. Ce rapport regroupe des résultats clés sur un mode facile à consulter pour le lecteur et se termine sur des résumés d'une page de toutes les évaluations examinées (ACDI, 2012).

³⁰Voir : www.norad.no/en/tools-publications/publications/search?search=®ion=&theme=&year=&type=annual-reports&serie=&publisher.

³¹Le rapport de la Conférence d'apprentissage sur Haïti de 2012 est disponible sur le Wikispace Apprentissage de la FICR : <http://ifrc-haiti-learning-conference.wikispaces.com/home>

Encadré

16

Le projet d'apprentissage des réponses d'urgence d'ACF

ACF International a mis en place en 2012 un projet d'apprentissage des réponses d'urgence pour aider l'organisation dans son développement de réponses plus opportunes, appropriées et efficaces aux urgences humanitaires du 21^e siècle. ACF a utilisé les évaluations externes de deux réponses d'urgence récentes (le tsunami asiatique en 2004 et le séisme en Haïti de 2010) pour identifier des thèmes clés autour desquels les réponses d'urgence sont construites et, dans un deuxième temps, pour identifier les bonnes pratiques et les principes qui ont prouvé leur plus grande efficacité lors de réponses ou qui ont fait l'objet de recommandations dans des évaluations. L'objectif était de faciliter ces leçons apprises pour alimenter de futures réponses. Les thèmes identifiés par la suite constituent la base d'un cadre pour de futures évaluations de manière à garantir un apprentissage cohérent et constant au sein de l'organisation.

Source : ACF Emergency Response Learning Project (2011). Note de briefing d'ACF-UK Evaluations, Learning & Accountability Unit.

De telles méta-approches et synthèses sont importantes pour extraire toute la valeur de coûteux processus d'évaluations. Elles aident à garantir que les résultats enregistrés à travers de nombreuses évaluations sont validés sans être cantonnés à un seul projet. Lorsque les résultats d'évaluations sont rassemblés et synthétisés, les utilisateurs potentiels de telles informations disposent de plus de données et d'éléments de preuve ; ils peuvent ainsi avoir une plus grande confiance dans les résultats qui leur sont présentés. Une plus grande homogénéité des résultats à travers les programmes renforce également leur crédibilité et la volonté de procéder à des changements qui s'en inspirent.

Encadré

17

Améliorer l'utilisation des résultats d'évaluation au sein du PAM

Le PAM a mobilisé des ressources importantes et réalisé des efforts significatifs pour améliorer l'impact qu'il obtient à partir de ses évaluations en se focalisant sur le processus post-évaluation de manière à augmenter l'accès à des éléments de preuve pertinents et opportuns tirés des évaluations ainsi que leur utilisation :

- Le PAM a réalisé des synthèses de revues et d'évaluations au niveau pays pour alimenter les processus de prises de décision au niveau pays comme le développement de nouvelles stratégies pays. La synthèse pays Éthiopie, par exemple, a puisé dans 16 évaluations et études pour présenter des leçons clés tirées d'expériences passées dans le même cadre que celui demandé par la stratégie pays prévisionnelle. Toutes les leçons renvoyaient aux rapports source pour faciliter une lecture plus complète. Les bureaux pays ont beaucoup apprécié.
- Le PAM a produit une synthèse des 4 évaluations stratégiques de différentes missions de son Plan stratégique 2008-2013. À partir de cette synthèse, d'autres produits sur mesure ont été réalisés, notamment un que les cadres dirigeants utiliseraient lors d'un atelier stratégique spécifique. Les documents de synthèse – en particulier pour des moments de prises de décision ponctuels – sont très prisés.
- Quatre documents thématiques portant sur les 10 meilleures leçons ont été préparés d'après un modèle fourni par ALNAP : sur le ciblage, les espèces et bons d'achats, le genre et les filets de sécurité. Des fiches d'évaluation ont été réalisées pour tous les rapports d'évaluation conclus avant 2011.
- Les évaluations de politiques alimentent le cycle de planification, de suivi et d'évaluation de politiques nouvelles et existantes. De plus, les évaluations des portefeuilles pays sont déjà fixées de manière à fournir des éléments de preuve pour la préparation des stratégies pays du PAM.
- Des leçons tirées de la série d'évaluations d'impact des cantines scolaires ont été présentées à une consultation interne sur le sujet et à une réunion technique internationale sur l'alimentation scolaire à l'aide de produits locaux. Un séminaire a été également organisé par le personnel à l'heure du déjeuner.
- Les évaluations sont accessibles dans la section Bibliothèque des évaluations du site Internet officiel du PAM. De multiples documents sont disponibles pour tirer des leçons d'évaluations en fonction des publics spécifiques.

Sources : Sally Burrows, WFP, communication personnelle, mars 2013 ; WFP, 2009 ; 2011.

Observations finales

Dans ce document, nous avons cherché à souligner des sujets clés renforçant la capacité de réaliser et d'utiliser – au sens le plus large du terme – l'évaluation des actions humanitaires (EAH). En nous appuyant sur la littérature publiée aussi bien que la littérature grise ou encore les entretiens et les ateliers avec des évaluateurs humanitaires, nous avons développé un cadre d'analyse des capacités d'évaluation dont le but est de faciliter l'auto-analyse des organisations et de structurer le débat autour des questions clés identifiées dans cette étude. Nous espérons que les organisations seront en mesure d'utiliser les résultats présentés dans ce cadre afin qu'elles puissent améliorer leurs capacités à financer, commanditer, soutenir, réaliser, communiquer et utiliser de façon conséquente les évaluations humanitaires.

Ce cadre est hiérarchique, ses entrées les plus importantes sont le leadership, la culture, la structure et les ressources qui apparaissent dans le Domaine de capacité n°1 (« Leadership, culture, structure et ressources »). Clarifier l'objectif, la demande et la stratégie sont également importants mais moins significatifs et, pour cela, apparaissent dans le Domaine de capacité n°2 (« Objectif, demande et stratégie »). Le Domaine de capacité 3 (« Processus et systèmes d'évaluation ») se focalise sur des processus et des systèmes qui, bien qu'utiles en soi, sont considérés comme moins à même d'apporter un changement fondamental par eux-mêmes, sans que des changements soient réalisés par ailleurs.

Il peut être plus facile de mettre en œuvre les recommandations du Domaine de capacités n°3 que celles des deux autres. De même, il se peut que des gains obtenus à un endroit amènent ailleurs un changement plus fondamental. Dans le même temps, les trois Domaines de capacités se renforcent mutuellement. Certaines victoires faciles provenant d'un Domaine de capacité « plus faible » peuvent aider à catalyser un intérêt pour l'évaluation qui peut promouvoir le changement dans des domaines de capacités plus exigeants. De petits changements dans un domaine peuvent remonter et « déborder » dans d'autres domaines de capacité^[32].

.....
³²Margie Buchanan-Smith, consultante indépendante, communication personnelle, juin 2013.

Il semble juste de noter qu'aucune approche uniforme ne peut être déduite de cette exploration des capacités de l'évaluation humanitaire. Comme d'autres l'ont remarqué, cela provient aussi du fait qu'augmenter l'impact des évaluations humanitaires passe également par le fait de trouver comment les résultats d'une évaluation peuvent poursuivre leur chemin et avoir des effets concrets au sein d'une organisation. Réaliser une évaluation de bonne qualité est seulement la première étape. Les leçons doivent ensuite être absorbées, digérées puis mises en pratique avant que l'apprentissage organisationnel ne soit achevé (Stoddard, 2005).

Nous avons alimenté le cadre d'analyse des capacités de l'évaluation humanitaire présenté dans ce document avec des études de cas et des indications glanées lors d'entretiens et d'ateliers mais aussi au moyen de la communauté de pratique de l'évaluation d'ALNAP. Pour certains domaines du cadre, il était plus difficile d'obtenir des exemples et des indications pratiques de la part des praticiens. Cette difficulté reflète peut-être des questions mal posées ou le fait que les expériences des évaluateurs et des praticiens dans ces domaines n'aient pas encore été enregistrées et analysées plus précisément.

Ce qui s'est avéré évident tout au long de ce processus, c'est que les évaluateurs doivent faire face à une énorme charge de travail. Aussi, nous avons grandement apprécié le temps qu'ils ont bien voulu consacrer à ce projet. En fait, l'énergie et la passion que nous avons rencontrées chez les évaluateurs du secteur humanitaire est stupéfiante. L'évaluation de l'action humanitaire reste une discipline relativement nouvelle et qui évolue constamment mais elle semble être entre de bonnes mains !

Le cadre suggéré dans cette étude est délibérément léger afin que les évaluateurs déjà bien occupés puissent y accéder de façon aisée et efficiente. Ce n'est pas un manuel présentant étape après étape comment mettre en œuvre le changement. Nous l'avons plutôt conçu pour motiver et inspirer, ainsi que pour servir de base de réflexion. Actuellement, un manuel « étape par étape » ne serait pas possible car il n'existe pas d'éléments de preuve sur le séquençage optimal des actions de renforcement des capacités pour les trois composantes du cadre. Toutefois, de nombreuses personnes interrogées pensent que le leadership demande plus d'attention au début. Les contributeurs de cette étude ont également vu comme très important le fait de sécuriser les ressources adéquates. Il semble cependant

probable qu'il n'existe pas une approche unique capable de fonctionner de façon universelle puisque les organisations du secteur sont très différentes en termes de taille, de culture et d'ambition.

De ce fait, il est important que les organisations conduisent leur propre analyse interne de manière à concevoir l'approche la plus appropriée à l'amélioration de l'utilisation de leurs évaluations. En guise de suivi de cette étude, un outil d'auto-évaluation sera conçu pour aider les organisations à réfléchir sur leurs processus d'évaluation, à enregistrer leurs pratiques en matière d'utilisation et d'adoption des évaluations mais également à identifier des domaines sur lesquels focaliser leurs futurs efforts^[33]. Une version pilote de l'outil a été testée par des praticiens et des évaluateurs humanitaires qui ont contribué à cette étude. Plusieurs organisations l'ont trouvé très utile lorsqu'il était confié à différentes sections et branches de l'organisation plutôt qu'au seul bureau des évaluations, les résultats étant alors discutés de manière conjointe. Le personnel de terrain et les évaluateurs peuvent percevoir et évaluer de façon sensiblement différente la capacité d'évaluation de leur propre organisation. L'outil peut donc aider à mettre en exergue certaines perceptions, justes ou fausses, qui gravitent souvent autour du travail d'évaluation et pourrait ainsi encourager les discussions visant à améliorer l'utilisation de l'évaluation.

Il existe de nombreux documents disponibles pour approfondir la lecture en matière de renforcement des capacités d'évaluation même s'ils ne sont pas toujours focalisés sur le secteur humanitaire. Beaucoup se concentrent sur le volet prestation et préconisent divers processus de changement de type structurel, personnel, technique et financier.

Il semble cependant probable qu'il n'existe pas une approche unique capable de fonctionner de façon universelle puisque les organisations du secteur sont très différentes en termes de taille, de culture et d'ambition.



Nous avons conçu ce document en adoptant une approche moins conventionnelle et en nous concentrant autant que possible sur la demande. Sans stimuler la demande d'évaluation, un changement structurel aura moins de chance de se produire.

Les membres de la communauté de l'évaluation humanitaire pourront continuer les débats grâce à la communauté de pratique de l'évaluation d'ALNAP (www.alnap.org/forum/evaluationcapacitiescop.aspx). Plus les expériences pour améliorer les capacités d'évaluation dans le secteur humanitaire seront nombreuses, plus cela amènera des indications riches et cela augmentera les connaissances quant à ce qui fonctionne. Il est important qu'un mécanisme permettant de partager cela continue d'exister.

Liste des contributeurs

Cette liste inclut toutes les personnes qui ont contribué à cette étude via un entretien ou une participation à une réunion de la communauté de pratique de l'évaluation d'ALNAP. Les exemples et les indications qu'ils ont partagés avec nous ont grandement enrichi ce travail et l'ont rendu plus proche des réalités des praticiens de l'évaluation.

Organisation	Contributeurs
ACF	Saul Guerrero, Ben Allen
American Red Cross	Dale Hill (aujourd'hui indépendant)
CDA	Dayna Brown
DARA	Riccardo Polastro
DEC	Annie Devonport
DFID	Jonathan Patrick
Groupe URD	François Grünwald, Véronique de Geoffroy
Independent	John Cosgrave, Margie Buchanan-Smith, Jock Baker, Tony Beck, Abhijit Bhattacharjee, Carlisle Levine, Monica Oliver
IRC	Anjuli Shivshanker
MSF	Sabine Kampmueller
Oxfam America	Silva Sedrakian
Oxfam GB	Vivien Walden
Oxfam Intermón	Luz Gómez Saavedra
Save the Children International	Dan Collison
Save the Children US	Hana Crowe
Sida	Joakim Molander (now Swedish Ministry of Foreign Affairs)
Tearfund	Catriona Dejean
UNHCR	Jeff Crisp
UNICEF	Rob McCouch (now UN OIOS), Erica Mattellone
UNRWA	Robert Stryk
WFP	Sally Burrows

Bibliographie

La bibliographie annotée est également disponible sur le site web d'ALNAP: <http://www.alnap.org/search/advanced.aspx?tag=543>.

ACDI (2012) *Annual report 2011–2012: CIDA learns – lessons from evaluations*.

ACF International (2011a) *Evaluation policy and guidelines: enhancing organisational practice through an integrated evaluations, learning and accountability framework*. London: ACF International.

ACF International (2011b) *Learning Review 2011*. London: ACF International.

ACF International (2013) *2012 Learning Review*. London: ACF International.

ACF-UK (2011) 'ACFIN emergency response learning project'. Briefing note. (internal document) London: Evaluations, Learning & Accountability Unit ACF-UK.

Allen, M. (2010) *Steeping the organization's tea: examining the relationship between evaluation use, organizational context, and evaluator characteristics*. Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Mandel School of Applied Social Sciences, Case Western Reserve University.

Apthorpe, R., Borton, J. and Wood, A. (eds) (2001) *Evaluating international humanitarian action: Reflections from Practitioners*. London: Zed Press.

ALNAP (2012) *The State of the Humanitarian System*. London: ALNAP/ODI.

ALNAP and Groupe URD (2009) *Participation handbook for humanitarian field workers: involving crisis-affected people in a humanitarian response*. Plaisians: Groupe URD.

Anderson, M., Brown, D. and Jean, I. (2012) *Time to listen: hearing people on the receiving end of international aid*. Cambridge, Massachusetts: CDA Collaborative Learning Projects.

Baker, J., Christoplos, I., Dahlgren, S., Frueh, S., Kliet, T., Ofir, Z. and Sandison P. (2007) *Peer review: evaluation function at the World Food Programme*. Sida.

Bamberger, M. (2012) *Introduction to mixed methods in impact evaluation*. Impact Evaluation Notes. Washington DC: InterAction.

Beck, T. (2002) 'Meta-evaluation' in *ALNAP Annual Review: improving performance through improved learning*. London: ALNAP/ODI.

Beck, T. (2004a) 'Meta-Evaluation' in *ALNAP Review of Humanitarian Action in 2003: Field level learning*. London: ALNAP/ODI.

Beck, T. (2004b) 'Learning and evaluation in humanitarian action', in *ALNAP Review of Humanitarian Action in 2003: Field level learning*. London: ALNAP/ODI.

Beck, T. (2006) 'Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies'. London: ALNAP/ODI.

Beck, T. and Buchanan-Smith, M. (2008) 'Joint evaluations coming of age? The quality and future scope of joint evaluations' in *ALNAP 7th Review of Humanitarian Action*. London: ALNAP/ODI.

Beck, T. and Buchanan-Smith, M. (2008) 'Meta-evaluation' in *ALNAP 7th Review of Humanitarian Action*. London: ALNAP/ODI.

Bonbright, D. (2012) 'Use of impact evaluation results'. Impact Evaluation Note 4. Washington, D.C.: InterAction.

Bourgeois, I., Hart, R., Townsend, S. and Gagné, M. (2011) 'Using hybrid models to support the development of organizational evaluation capacity: A case narrative', *Evaluation and Program Planning* 34: 228–235.

Boyle, R. and Lemaire, D. (eds) (1999) 'Building effective evaluation capacity: lessons from practice'. Comparative Policy Evaluation 8. New Brunswick: Transaction Publishers (www.alnap.org/resource/8501.aspx).

Buchanan-Smith, M. with Scriven, K. (2011) *Leadership in action: leading effectively in humanitarian operations*. London: ALNAP/ODI.

- Catley, A., Burns, J., Abede, D. and Suji, O. (2008) *Participatory impact assessment: a guide for practitioners*. Medford, MD: Feinstein International Center, Tufts University.
- Carlsson, J., Koehlin, G. and Ekborn, A. (1994) *The political economy of evaluation: international aid agencies and the effectiveness of aid*. London: Macmillan Press.
- Chelimsky, E. (1997) 'The coming transformations in evaluation', in Chelimsky E. and Shadish, W. (eds), *Evaluation for the 21st century: A Handbook*. Thousand Oaks, CA: Sage.
- Clarke, P. and Ramalingam, B. (2008) *Organisational change in the humanitarian sector*. London: ALNAP/ODI.
- Cosgrave, J. (2008) *Responding to earthquakes: learning from earthquake relief and recovery operations*. ALNAP Lessons Paper. London: ALNAP/ODI.
- Cosgrave, J. and Buchanan-Smith, M. (2013) *Evaluating humanitarian action: ALNAP pilot guide*. London: ALNAP/ODI.
- Cosgrave, J., Ramalingam, B. and Beck, T. (2009) *Real-time evaluations of humanitarian action – an ALNAP guide*. Pilot version. London: ALNAP/ODI.
- Cosgrave, J., Polastro, R. and van Eekelen, W. (2012) *Evaluation of the Consortium of British Humanitarian Agencies (CBHA) pilot*.
- Crisp, J. (2000) 'Thinking outside the box: evaluation and humanitarian action', in *Forced Migration Review* 8: 4-7.
- DAC/UNEG (2007) 'Framework for professional peer reviews'. DAC/UNEG Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations. New York: UNEG.
- Darcy, J. (2005) 'Acts of faith? Thoughts on the effectiveness of humanitarian action'. Paper prepared for the Social Science Research Council seminar series, *The transformation of humanitarian action*, 12 April, New York. London: Overseas Development Institute.
- DEC (2013) 'DEC approach to learning'. Paper 7, February. (internal document) London: DEC.
- DFID (2011a) 'How to note: writing a business case'. DFID Practice Paper.
- DFID (2011b) 'Department for International Development organisation chart'.
- DFID (2013) *International development evaluation policy*. London: Department for International Development.
- Drew, R. (2011) *Synthesis study of DFID's strategic evaluations 2005–2010*. Report produced for the Independent Commission for Aid Impact.
- ECB (2007) *Good enough guide: impact measurement and accountability in emergencies*. Warwickshire: Practical Action Publishing.
- ECB (2011) 'What we know about joint evaluations of humanitarian action: learning from NGO experiences'. Version 6, April.
- ECB - AIM Standing Team (2011) *Case study on Accountability and Impact Measurement Advisory Group*.
- Engel, P., Carlsson, C. and van Zee, A. (2003) *Making evaluation results count: internalising evidence by learning*. Policy Management Brief 16.
- FAO (2012) *Peer Evaluation: The Evaluation Function of the Food and Agriculture Organization*. Final Report. Washington, DC: Global Facility Evaluation Office.
- Fearon, J. (2005) 'Measuring humanitarian impact'. Discussion paper for the SSRC seminar series, *The Transformation of Humanitarian Action*, 12 April, New York.
- Few, R., McAvoy, D., Tarazona, M. and Walden, V. (2013) *Contribution to Change: An approach to evaluating the role of intervention in disaster recovery*. Oxford: Oxfam.
- Feinstein, O. (2002) 'Use of evaluations and the evaluation of their use', *Evaluation* 8(4): 433–439.
- Foresti, M. (2007) 'A comparative study of evaluation policies and practices in development agencies'. Série Ex Post : notes méthodologiques 1, December. London: ODI.
- Forss, K., Krause, S.-E., Taut, S. and Tenden, E. (2006) 'Chasing a ghost? An essay on participatory evaluation and capacity development', *Evaluation* 1(6): 129.
- Frerks, G. and Hilhorst, D. (2002) 'Evaluation of humanitarian assistance in emergency situations'. Working Paper 56. New Issues in Refugee Research, UNHCR/EPAU.

Furubo, J.-E., Rist, R. and Sandahl, R. (eds) (2002) *International Atlas of Evaluation*. New Brunswick: Transaction Publishers.

Government Office for Science (2011a) 'Science & engineering assurance review of the Department for International Development'. URN 11/1260.

Government Office for Science (2011b) 'DIFD science & engineering assurance review: annex A – department background'. URN 11/1260.

Grasso, P. (2003) 'What makes an evaluation useful? Reflections from experience in large organizations', *American Journal of Evaluation* 24(4): 507–514.

Griekspoor A. and Sondorp, E. (2001) 'Enhancing the quality of humanitarian assistance: taking stock and future initiatives', in *Prehospital and Disaster Medicine* 16(4): 209–215.

Groupe URD (2009) *The Quality Compass companion book*. Plaisians : Groupe URD.

Groupe URD (2010) *Étude en temps réel de la gestion de la crise en Haïti après le séisme du 12 janvier 2010*. Plaisians : Groupe URD.

Guba, E. and Lincoln, Y. (1989) *Fourth Generation Evaluation*. Newbury Pak, CA: Sage.

Guerrero, S. (2012) 'Learning to learn: harnessing external evaluations to build on strengths, enhance effectiveness'. ACF Blog, 14 August. London: Action Against Hunger.

Guerrero, S. (2013) 'The needle in the haystack: the role of external evaluations in identifying and promoting good practices within NGOs'. Presentation prepared for the 28th ALNAP Meeting, Washington DC, 5-7 March.

Guerrero, S., Woodhead, S. and Hounjet, M. (2013) *On the right track: a brief review of monitoring and evaluation in the humanitarian sector*.

Hallam, A. (1998) 'Evaluating humanitarian assistance programmes in complex emergencies'. Relief and Rehabilitation Network, Good Practice Review 7. London: ODI.

Hallam, A. (2011) 'Harnessing the power of evaluation in humanitarian action: an initiative to improve understanding and use of evaluation'. ALNAP Working Paper. London: ALNAP/ODI.

Heider, C. (2011) 'Conceptual framework for developing evaluation capacities: building on good practice', in Rist, R., Boily, M.-H. and Martin, F. (eds), *Influencing Change: Evaluation and Capacity Building: 85–107*. Washington DC: World Bank.

House, E. (1983) 'Assumptions underlying evaluation models', in Madaus, G. F., Scriven, M. and Stufflebeam, D. L. (eds), *Evaluation Models*. Boston: Kluwer-Nijhoff.

Hovland, I. (2003) 'Knowledge management and organisational learning, an international development perspective: an annotated bibliography'. ODI Working Paper 224. London: ODI.

IASC (2008) *Inter-agency RTE of the humanitarian response to cyclone Nargis in Myanmar*. Geneva and New York: IASC.

IASC (2010) *Inter-agency real-time evaluation procedures and methodologies*. Geneva and New York: IASC.

IASC (2011a) 'Inter-agency real-time evaluation of the humanitarian response to 2010 floods in Pakistan'. Geneva and New York: IASC.

IASC (2011b) 'Inter-agency real-time evaluation of the humanitarian response to the earthquake in Haiti: 20 months after'. Geneva and New York: IASC.

IASC (2012) *IASC transformative agenda – 2012*. Geneva and New York: IASC.

ICAI (2011a) 'Independent Commission for Aid Impact – Work Plan'.

ICAI (2011b) 'Independent Commission for Aid Impact: report to the House of Commons International Development Committee on ICAI's Shadow Period: November 2010 – May 2011'.

ICAI (2012) 'Transparency, impact and Value for Money – The work of the ICAI and how it impacts INGOs'. Speech by Chief Commissioner at the Crowe Clark Whitehill INGO conference on 3 February 2012.

- IFRC (2011a) *IFRC framework of evaluation*. IFRC Secretariat: Planning and Evaluation Department.
- IFRC (2011b) *Project/programme monitoring and evaluation (M&E): guide*. Geneva: IFRC.
- IFRC (2011c) *Revised plan 2011: disaster services – executive summary*. Geneva: IFRC.
- IFRC (2013) 'Learning from the Haiti operation: from a large-scale operation experience to a global and institutionalized knowledge'. Learning Concept Paper for Haiti. (internal document) Port-au-Prince: IFRC.
- Innovation Network (2013) *State of evaluation 2012: evaluation practice and capacity in the nonprofit sector*.
- Jones, N., Jones, H., Steer, L. and Datta, A. (2009) *Improving impact evaluation production and use*. Working Paper 300. London: ODI.
- Jones, H. and Mendizabal, E. (2010) *Strengthening learning from research and evaluation: going with the grain*. London: ODI.
- Kellogg Foundation (2007) *Designing Initiative Evaluation*. Battle Creek, MI: WK Kellogg Foundation.
- Lawrenz, F., Gullickson, A. and Toal, S. (2007) 'Dissemination: handmaiden to evaluation use', in *American Journal of Evaluation* 28(3): 275–89.
- Levine, C. and Chapoy, C. (2013) *Building on what works: Commitment to Evaluation (C2E) Indicator*.
- Lincoln, Y. (1991) 'The arts and sciences of program evaluation', in *Evaluation Practice* 12(1): 1.
- Lindahl, C. (1998) 'The management of disaster relief evaluation - Lessons from a Sida Evaluation of the Complex Emergency in Cambodia', in Sida Department for Evaluation and Internal Audit, Sida Studies in Evaluation.
- Lindahl, C. (2001) 'Evaluating SIDA's complex emergency assistance in Cambodia: conflicting perceptions', in Wood, Apthorpe and Borton (eds), *Evaluating International Humanitarian Action*: 59. London: Zed Books.
- Liverani, A. and Lundgren, H. (2007) 'Evaluation systems in development aid agencies: an analysis of DAC peer reviews 1996–2004', in *Evaluation* 13(2): 242.
- Mayne, J. (2008) 'Building an evaluative culture for effective evaluation and results management'. Institutional Learning and Change Initiative (ILAC) Brief 20, November.
- Mayne J. and Rist, R. (2006) 'Studies are not enough: the necessary transformation of evaluation', in *Canadian Journal of Program Evaluation* 21(3): 93–120.
- Mackay, K. (2009) 'Building monitoring and evaluation systems to improve government performance' in *Country-led monitoring and evaluation systems - Better evidence, better policies, better development results*. Segone, M. (ed.) New York: UNICEF. 169-187.
- Mebrahtu, E., Pratt, B. and Lönnqvist, L. (2007) *Rethinking Monitoring and Evaluation: Challenges and Prospects in the Changing Global Aid Environment*. Oxford: INTRAC.
- McNulty, J. (2012) 'Symbolic uses of evaluation in the international aid sector: arguments for critical reflection', in *Evidence & Policy: a Journal of Research, Debate and Practice*, 8(4): 495-509.
- Minnett, A. M. (1999) 'Internal evaluation in a self-reflexive organization: One nonprofit agency's model', *Evaluation and Program Planning* (22): 353-362.
- Molund, S. and Schill, G. (2007) *Looking back, moving forward: Sida Evaluation Manual*. Sida.
- Morras Imas, L. and Rist, R. (2009) *The road to results: designing and conducting effective development evaluations*. Washington DC: The World Bank.
- Nelson, J. (2012) 'Actionable measurement at the Gates Foundation', Blog from 6 September 2012.
- NGO and Humanitarian Reform Project (2010) 'Good practices in humanitarian assistance: Zimbabwe'. Good Practice Paper Series.
- OECD (2006) 'The Challenge of Capacity Development: Working Towards Good Practice'. DAC Guidelines and Reference Series.
- OECD-DAC (2001) 'Evaluation feedback for effective learning and accountability'. Evaluation and Aid Effectiveness 5. Paris: OECD.
- OECD-DAC (2002) *Glossary of Key Terms in Evaluation and Results-Based Management*. Paris: OECD/DAC.

- Oliver, M. (2007) *CARE's humanitarian operations: review of CARE's use of evaluations and after action reviews in decision-making*. Georgia State University.
- Oliver, M. (2009) 'Meta-evaluation as a means of examining evaluation influence', in *Journal of MultiDisciplinary Evaluation* 6(11): 32–37.
- O'Neil, G. (2012) '7 new ways to communicate findings'. Presentation at the European Evaluation Society Conference in Helsinki, Finland.
- Oral History Project Team (2006) 'The oral history of evaluation, Part IV – the professional evolution of Carol H. Weiss', in *American Journal of Evaluation* 27(4): 478.
- Owen, J. and Lambert, F. (1995). 'Roles for evaluation in learning organisations', in *Evaluation*. 1 (2), 259-273.
- Oxfam GB (2012) *Project effectiveness reviews: summary of 2011/12 findings and lessons learned*. Oxford: Oxfam.
- Oxfam GB (n.d.) *How are effectiveness reviews carried out?* Oxford: Oxfam.
- Pankaj, V., Welsh, M. and Ostenso, L. (2011) 'Participatory analysis: expanding stakeholder involvement in evaluation'. Innovation Network.
- Patton, M.Q. (1990) 'The evaluator's responsibility for utilisation', in Alkin, M. (ed.) *Debates on Evaluation*. Newbury Park, CA: SAGE.
- Patton, M.Q. (1997) *Utilization-focused evaluation: the new century text*. 3rd ed. Thousand Oaks, CA: Sage.
- Patton M.Q. (2001) 'Evaluation, knowledge management, best practices, and high. quality lessons learned', in *American Journal of Evaluation* 22(3).
- Patton, M.Q. (2010a) *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York: Guilford Press.
- Patton, M.Q. (2010b) 'Incomplete successes', in *Canadian Journal of Program Evaluation* 25(3), Special issue.
- Pawson, R. and Tilley, N. (2004) *Realist evaluation*. Mount Torrens: Community Matters.
- Perrin, B. (2013) 'Evaluation of Payment by Results (PBR): Current Approaches, Future Needs'. Working Paper 39. London: DFID.
- Picciotto, R. (2007) 'The new environment for development evaluation', in *American Journal of Evaluation* 28(4): 521.
- Pratt, B. (2007) 'Rethinking monitoring and evaluation', in *Ontrac* 37 (September): 1.
- Preskill, H. and Torres, R. (1999) *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.
- Preskill, H. and Boyle, S. (2008) 'A multidisciplinary model of evaluation capacity building', in *American Journal of Evaluation* 29: 443–459.
- Proudlock, K. (2009) 'Strengthening evaluation capacity in the humanitarian sector: a summary of key issues' (author's unpublished notes).
- Proudlock, K., Ramalingam, B. with Sandison, P. (2009) 'Improving humanitarian impact assessment: bridging theory and practice', in *ALNAP 8th Review of Humanitarian Action*. London: ALNAP/ODI.
- Quesnel, J. S. (2006) 'The importance of evaluation associations and networks', in Segone, M. (ed), *New Trends in Development Evaluation*, Evaluation Working Papers, 5: 17-25. UNICEF, CEE/CIS, IPEN.
- Ramalingam, B. (2005) 'Implementing knowledge strategies: lessons from international development agencies'. ODI Working Paper 244. London: Overseas Development Institute.
- Ramalingam, B. (2006) *Tools for knowledge and learning: a guide for development and humanitarian organisations*. London: Overseas Development Institute.
- Ramalingam, B. (2011) 'Learning how to learn: eight lessons for impact evaluations that make a difference'. ODI Background Notes (April). London: Overseas Development Institute.
- Reed, E. and Morariu, J. (2010) 'State of evaluation: evaluation practice and capacity in the nonprofit sector'. Innovation Network.

- Results Matter (2013) 'Humanitarian Genome Project: end of project evaluation, June 2013' (internal document).
- Rey, F. and Urgoiti, A. (2005) 'Manual de gestión del ciclo del proyecto en acción humanitaria'. Fundación la Caixa.
- Ridde, V. and Sahibullah, S. (2005) 'Evaluation capacity building and humanitarian organization', *Journal of MultiDisciplinary Evaluation (JMDE)* 3: 78–112.
- Riddell, R. (2009) 'The quality of DFID's evaluation reports and assurance systems: a report for ICAI based on the quality review' (internal document) London: IACDI.
- Rist, R. (1990) *Program evaluation and the management of government: patterns and prospects across eight nations*. New Brunswick: Transaction Publishers.
- Rist, R., Boily, M.-H. and Martin, F. (2011) 'Introduction. Evaluation capacity building: a conceptual framework', in Rist, R., Boily, M.-H. and Martin, F. (eds), *Influencing Change: Evaluation and Capacity Building*: 1–11. Washington, DC: World Bank.
- Rist, R., Boily, M.-H. and Martin, F. (2013) *Development evaluation in times of turbulence: dealing with crises that endanger our future*. Washington, DC: World Bank.
- Rogers, P. (2009) 'Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation in designing impact evaluations: different perspectives'. 3ie Working Paper 4. London: 3ie.
- Rossi, P. and Freeman, H. (1993) *Evaluation: a systematic approach*. 5th ed. Newbury Park, CA: Sage.
- Rugh, J. and Segone, M. (eds) (2013) 'Voluntary Organizations for Professional Evaluation (VOPEs) learning from Africa, Americas, Asia, Australasia, Europe and Middle East'. EvalPartners, UNICEF, IOCE.
- Sandison, P. (2006) 'The utilisation of evaluations', *ALNAP Review of Humanitarian Action: Evaluation Utilisation*. London: ALNAP/ODI.
- SCHR (2010) *SCHR peer review on accountability to disaster-affected populations*. Geneva: SCHR.
- Scriven, M. (1991) *Evaluation thesaurus*. 4th ed. Newbury Park, CA: Sage.
- Segone M. and Ocampo, A. (eds) (2007) *Creating and developing evaluation organizations - lessons learned from Africa, Americas, Asia, Australasia and Europe*. Lima: IOCE, desco and UNICEF.
- Spilsbury, M.J., Perch, C., Norgbey, S., Rauniyar, G. and Battaglino, C. (2007) 'Lessons learned from evaluation: a platform for sharing knowledge'. Special Study Paper Number 2. Nairobi: Evaluation and Oversight Unit, United Nations Environment Programme.
- Start, D. and Hovland, I. (2004) 'Tools for policy impact: a handbook for researchers'. Toolkit. London: ODI.
- Stern, E. et al. (2012) 'Broadening the range of designs and methods for impact evaluations'. DFID Working Paper 38.
- Stoddard, A. (2005) *Review of UNICEF's evaluation of humanitarian action*. New York: United Nations Children's Fund.
- Stufflebeam, D. (2000) 'Lessons in contracting for evaluation', in *American Journal of Evaluation* 21: 293–314.
- Taut, S. (2007) 'Defining evaluation capacity building: utility considerations', in *American Journal of Evaluation* 28(1).
- Tavistock Institute (2003) *The evaluation of socio-economic development. The Guide*.
- Tearfund (2013) 'Impact, monitoring and evaluation planning and tracking tool' (internal document) Teddington: Tearfund.
- Torres, R.T. (1991) 'Improving the quality of internal evaluation: the evaluator as consultant-mediator', in *Evaluation and Program Planning* 14: 189–198.
- UNEG (2005) *Norms for evaluation in the UN system*. New York: UNEG.
- UNEG (2008) *UNEG Code of Conduct for Evaluation in the UN system*. New York: UNEG.
- UNEG (2010) *UNEG good practice guidelines for follow up to evaluations*. New York: UNEG.
- UNEG (2011) *UNEG framework for professional peer reviews of the evaluation function of UN organizations*. Reference Document. New York: UNEG.

- UNEG (2012a) 'National evaluation capacities'. Proceedings from the 2nd International Conference, 12–14 September 2011. UNEG Members' Publications.
- UNEG (2012b) *Evaluation capacity in the UN System*. New York: UNEG.
- UNEG (2012c) 'National evaluation capacity development: practical tips on how to strengthen National Evaluation Systems'. New York: UNEG.
- UNEG (2012d) 'Update note on peer reviews of evaluation in UN organizations'. UNEG-DAC Peer Reviews. New York: UNEG.
- UNEG (2013) 'Lessons learned study of peer reviews of UNEG evaluation functions'. UNEG-DAC Peer Reviews. New York: UNEG.
- UNEG/MFA (UNEG and Evaluation Department, Ministry of Foreign Affairs of Denmark) (2010) 'Peer review of the evaluation function at UNIDO'. New York: UNEG.
- UNICEF (2006) 'New trends in development evaluation'. Segone, M. (ed.). Evaluation Working Papers 5. Geneva: UNICEF CEE/CIS, Regional Office for Central and Eastern Europe and the CIS Countries.
- UNICEF (2010a) *Bridging the gap: the role of monitoring and evaluation in evidence-based policy-making*. New York: UNICEF.
- UNICEF (2010b) 'Evaluation of the UNICEF education programme in Timor Leste: 2003-2009'. New York: UNICEF.
- UNICEF (2010c) 'Global Evaluation Reports Oversight System (GEROS)'. New York: UNICEF.
- UNICEF (2012a) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): Cluster Analysis Summary Matrix' (internal document) New York: UNICEF (www.alnap.org/resource/8268.aspx).
- UNICEF (2012b) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): country case study selection – technical note' (internal document) New York: UNICEF.
- UNICEF (2012c) 'Evaluation of UNICEF's Cluster Lead Agency Role in Humanitarian Action (CLARE): terms of reference'. Final Draft, 22 October.
- UNICEF (2013) 'Thematic synthesis report on evaluation of humanitarian action'. Discussion paper submitted to the Executive Board for UNICEF's Annual Session 2013.
- UN General Assembly (2012) Resolution 67/226 on the Quadrennial Comprehensive Policy Review, adopted 21 December 2012.
- USAID (2009) 'Trends in international development evaluation theory, policy and practices'.
- Vogel, I. (2012) 'Review of the use of "theory of change" in international development. London: DFID.
- Wagenaar, S. (2009) 'Peer review on partnership in humanitarian aid: an approach for mutual learning and cooperation between organisations (North and South)'. Wageningen: PSO.
- Webster, M. and Walker, P. (2009) *One for all and all for one: Intra-organizational Dynamics in humanitarian action*. Medford, MA: Feinstein International Center.
- Weiss, C. (1983a) 'Ideology, interests, and information: the basis of policy positions', in Callahan, D. and Jennings, B. (1983) *Ethics, the Social Sciences, and Policy Analysis*. New York: Plenum.
- Weiss, C. (1983b) 'The stakeholder approach to evaluation: origins and promise', in Anthony Bryk (ed.), *New Directions for Programme Evaluation: Stakeholder-Based Evaluation*: 3-14. New Direction for Programme Evaluation Series 17. San Francisco: Jossey Bass.
- Weiss, C. (1990) 'Evaluation for decisions: is anybody there? Does anybody care?', in M. Alkin (ed.), *Debates on Evaluation*: 171–184. Newbury Park, CA: Sage.
- Weiss, C. (1998) *Evaluation – Methods for studying programs and policies* (2nd edition). NJ: Prentice Hall.
- WFP (2008a) *Evaluation of WFP's capacity development policy and operations*. Report. Rome: Office of Evaluation, World Food Programme.
- WFP (2008b) *WFP evaluation policy*.
- WFP (2009) *Closing the learning loop – harvesting lessons from evaluations: report of phase I*. Rome: Office of Evaluation, World Food Programme.

WFP-OE (2012) *Annual evaluation report: measuring results, sharing lessons*.

Wiles, P. (2005) 'Meta-evaluation' in *ALNAP Review of Humanitarian Action in 2004: capacity building*. London: ALNAP/ODI.

Wood, A., Apthorpe, R. and Borton, J. (eds) (2001) *Evaluating international humanitarian action: reflections from practitioners*. London: Zed Books.

World Bank (2009) 'Making smart policy: using impact evaluation for policy making'. *Doing impact evaluation* 14.

WB-IEG (2009) *Institutionalising impact evaluation within the framework of a monitoring and evaluation system*. Washington DC: World Bank Independent Evaluation Group.

York, N. (2012) 'Results, value for money and evaluation – a view from DFID'. Presentation delivered at the Development evaluation days, organized by AECID. Madrid 7 December 2012.

Young, J. and Mendizabal, E. (2009) 'Helping researchers become policy entrepreneurs'. ODI Briefing Paper 53. London: ODI.

Young, J. and Court, J. (2004) 'Bridging research and policy in international development: an analytical and practical framework'. RAPID Briefing Paper. London: ODI.

Autres publications d'ALNAP

Aussi disponible au www.alnap.org/publications

**Evaluating Humanitarian Action:
ALNAP pilot guide**

**Harnessing the power of evaluation in humanitarian action:
An initiative to improve understanding and use of evaluation**

**Improving humanitarian impact assessment:
Bridging theory and practice**

**Joint evaluations coming of age?
The quality and future scope of joint evaluations**

**Evaluating humanitarian action
using the OECD-DAC criteria**

The utilisation of evaluations

ALNAP
Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ
Royaume-Uni
Tél : + 44 (0)20 3327 6578
Fax : + 44 (0)20 7922 0399
Email: alnap@alnap.org

Remerciements à USAID pour le financement
de cette Étude ALNAP.

