

Chapitre 4

Conduire une évaluation dans les situations de conflit ou de fragilité

Le chapitre 4 examine le processus d'évaluation. Il commence par la phase de démarrage et s'interroge ensuite sur les moyens de définir des théories du changement et la logique de mise en œuvre qui sous-tend l'activité évaluée. Il aborde ensuite l'insuffisance des données de référence et définit comment et où obtenir des données de manière à combler les lacunes constatées. Il analyse les critères que les évaluateurs devraient utiliser, et plus particulièrement la pertinence, la durabilité, l'efficacité, l'efficience et l'impact des activités du CAD. Il décrit ensuite comment mener une évaluation à terme. Il examine la façon de dégager des conclusions et de formuler des recommandations, et se penche sur les rapports établis par les évaluateurs. Il s'intéresse ensuite à la communication des résultats (positifs ou négatifs) de l'évaluation aux parties intéressées et à la diffusion des enseignements dégagés. Enfin, pour boucler la boucle, il insiste sur la nécessité d'intégrer les conclusions de l'évaluation à la conception et à la gestion des programmes.

Toute évaluation doit comporter certaines étapes fondamentales. Les évaluateurs et les responsables (et les autres participants au processus concret d'évaluation) ne sont pas tenus de les suivre dans l'ordre présenté ici, mais doivent néanmoins les garder à l'esprit.

Étapes fondamentales du processus d'évaluation	
✓	Prévoir une phase de démarrage
✓	Définir la logique d'intervention et la théorie du changement
✓	Remédier au problème des données de référence manquantes et aux autres lacunes
✓	Collecter des données
✓	Évaluer l'activité au moyen de différents critères
✓	Dégager des conclusions et formuler des recommandations
✓	Établir des comptes rendus
✓	Veiller à la qualité
✓	Réagir à l'évaluation
✓	Réponse de la direction
✓	Diffusion des constats
✓	Intégration à la programmation et à l'apprentissage

Prévoir une phase de démarrage

Étant donné la difficulté à conduire des évaluations de l'action des donateurs dans les situations de fragilité ou de conflit et des faiblesses sous-jacentes de la programmation, une phase de démarrage permettra éventuellement de recenser les problèmes à régler avant de procéder à l'évaluation. Elle peut couvrir un éventail d'activités allant des plus simples aux plus compliquées. Elle comporte généralement une étude théorique, une étude documentaire et la publication d'un rapport de démarrage, mais des travaux préliminaires sur le terrain peuvent également s'imposer. Cette phase peut être utile pour l'évaluation ou la conduite d'analyses des conflits. Elle peut également servir à définir les contours précis de l'évaluation, à effectuer une analyse des politiques et du portefeuille des donateurs ou des politiques nationales ou, le cas échéant, à conduire des études thématiques.

Les évaluateurs présentent les résultats de la phase de démarrage dans un rapport susceptible de servir d'assise à des discussions portant sur la disponibilité des données, l'évaluabilité, et la possibilité d'appliquer les stratégies de collecte de données prévues. Ces rapports sont également des instruments utiles pour adapter les méthodes et peaufiner l'approche retenue afin de traiter les questions clés spécifiées dans le cahier des charges. (Voir plus loin la section intitulée « Rapports »).

Définir et évaluer la théorie du changement et la logique d'intervention

Une théorie du changement est un ensemble d'hypothèses sur la façon dont le changement intervient. Elle explique pourquoi et comment certaines mesures sont censées produire les changements souhaités dans un contexte donné et à un moment précis (Weiss, 1995 ; Church et Rogers, 2006). Cette expression est utilisée pour décrire les liens entre le contexte, les ressources de l'intervention, la stratégie de mise en œuvre, les produits et effets directs visés. La logique d'intervention, également dénommée logique du programme, est un terme qui décrit les raisons pour lesquelles une activité évolue de telle ou telle manière, généralement au niveau du projet. Elle est étroitement associée à la théorie du changement. L'évaluation peut révéler si la réussite, ou l'échec, est imputable à la conception et au principe du programme ou à son exécution. La vérification des théories du changement et de la logique d'intervention est l'une des principales contributions que peut apporter l'évaluation à la recherche et à l'apprentissage en général.

Dans les situations de conflit ou de fragilité, il est parfois particulièrement important que les évaluateurs définissent les théories du changement dans la mesure où elles sont souvent implicites, non vérifiées et non avérées. Au Kosovo, par exemple, la communauté internationale est intervenue pendant plusieurs années en se fondant sur l'hypothèse (théorie du changement) qu'il était possible d'instaurer la paix en améliorant les relations entre les deux principales parties au conflit. Elle a ainsi financé de nombreux programmes visant à favoriser le dialogue, les échanges, l'interaction entre les jeunes, les organisations de femmes, etc. Tous visaient à créer des liens entre les communautés. Or, une étude a observé qu'en l'absence de travaux intracommunautaires visant à remédier aux fractures internes et à créer un leadership plus responsable, les interventions intercommunautaires avaient peu d'effets. Les activités reposaient sur une théorie du changement incomplète. L'étude a également constaté que les relations intercommunautaires ne concouraient ni à prévenir la violence, ni à renforcer la résistance collective à la violence en 2004. L'établissement d'un lien entre le capital social (réseaux intra-ethniques de confiance et de réciprocité) et un leadership monocommunautaire, motivé par le souhait stratégique d'obtenir l'indépendance du Kosovo, a donné lieu à une action plus efficace (Chigas et al., 2006).

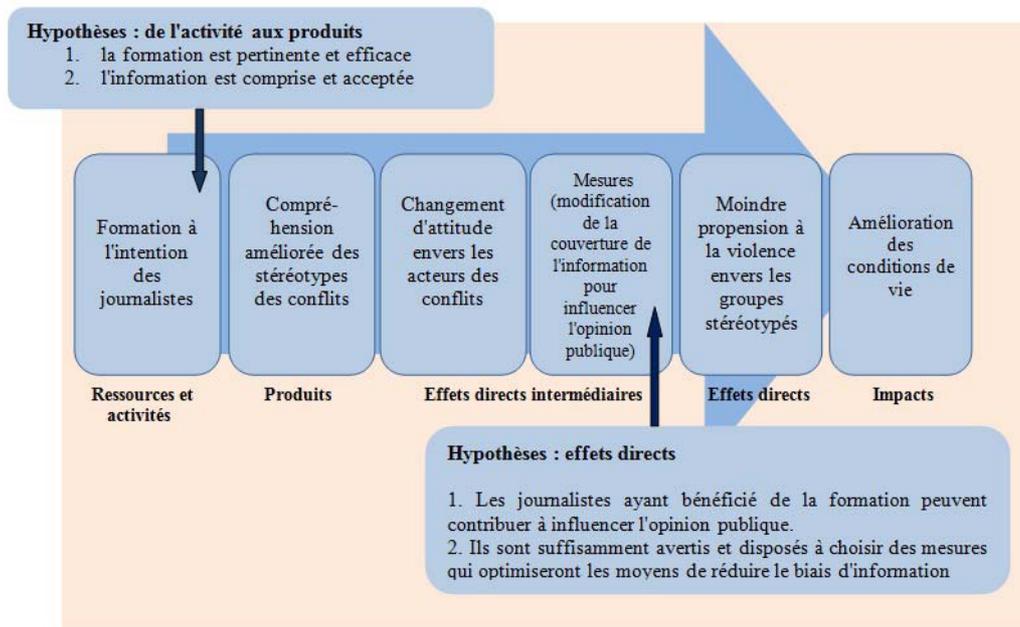
Causalité : entre théorie et effets directs

Une théorie du changement peut revêtir différentes formes à différents niveaux. Au niveau des pays et au niveau des conflits, ces théories peuvent être très générales – par exemple, le renforcement des capacités de l'État permettra d'améliorer la gouvernance et, par conséquent, de résorber le conflit et la violence. Dans cette optique globale, les acteurs du développement peuvent intervenir par différents moyens (appui budgétaire ou soutien à l'amélioration des capacités de certains ministères) ou dans différents secteurs, comme les services sociaux ou le financement et la planification.

Les évaluations de projets ou d'un programme précis formuleraient probablement des théories du changement plus directement causales, associant les moyens et les activités aux effets directs souhaités au microniveau et à la dynamique de paix en général. Ainsi, dans le cadre de l'évaluation du Soudan du Sud (encadré 3.6), l'équipe a identifié des théories générales du changement dans l'ensemble du portefeuille d'interventions considéré. Elle n'a pas affiné l'analyse au niveau particulier de la logique d'intervention ni examiné les différents projets. Elle a en revanche regroupé les multiples activités des différents donateurs sous des théories globales du changement, et vérifié ensuite chacune d'elles.

Les documents relatifs aux programmes ne présentent pas toujours en termes de causalité explicites la façon dont un programme entend obtenir les produits, les effets directs et les impacts visés. Des théories du changement précises et clairement énoncées sont nécessaires à une programmation efficace, et doivent faire l'objet d'une évaluation. Les hypothèses qui sous-tendent les théories, surtout, doivent être évaluées. En définissant par quels moyens une intervention était censée concourir à une paix durable ou remédier à des facteurs de conflit ou de fragilité, les évaluateurs utilisent la théorie du changement pour évaluer la pertinence, l'efficacité et l'impact des activités. Le programme fictif de journalisme pour la paix anti préjugés décrit à l'annexe C et illustré au graphique 4.1 offre un exemple de cette réflexion.

Graphique 4.1. **Hypothèses sous-jacentes à la théorie du changement dans un programme fictif de journalisme pour la paix**



Source : Texte de OCDE (2008b) ; schéma adapté de Gaarder, 3IE (2011).

Le programme pourrait être compris comme se fondant sur une théorie du changement en vertu de laquelle il assurera la formation de journalistes et améliorera leur connaissance de la dynamique du conflit, diminuant ainsi les biais d'information et favorisant un apaisement des tensions et de la violence entre les groupes en conflit. Une évaluation de ce programme vérifierait le bien-fondé de la théorie sous-jacente selon laquelle l'approfondissement des connaissances réduit les biais d'information et la théorie secondaire selon laquelle l'amélioration de l'information atténue la tendance à recourir à la violence dans la communauté au sens large. Il s'agirait essentiellement d'analyser en quoi les ateliers, les activités de sensibilisation et l'amélioration des compétences modifient concrètement la couverture des conflits et en quoi ils influencent leur dynamique fondamentale. Le programme pourrait analyser le style utilisé dans les reportages avant et après la formation. Il pourrait également observer l'état d'esprit du public et, concomitamment, les activités conduites dans le cadre du programme afin de

vérifier si elles obtenaient les résultats visés ou si des obstacles inattendus avaient surgi. Il pourrait par exemple s'avérer que les journalistes n'influent guère sur la phraséologie vindicative employée et que les rédacteurs choisissent de recourir à un style « haut en couleurs » pour stimuler les ventes, alimentant ainsi les images stéréotypées. Cette constatation montrerait que la « théorie » consistant à former les journalistes pour influencer sur l'opinion publique était sans fondement.

D'où viennent les théories du changement ?

Pour identifier et évaluer les théories du changement, il convient de déterminer d'où elles viennent. Sont-elles empiriques, fondées sur de solides données d'observation issues de programmes antérieurs ? Sont-elles le fruit de l'expérience personnelle et professionnelle des concepteurs du programme, ou de celle des parties prenantes et des bénéficiaires consultés durant son élaboration ? Représentent-elles sur des études ?

Dans les cas où les hypothèses de travail ne sont pas explicites, il conviendra de les définir au moyen d'un examen documentaire ou d'entretiens avec le personnel d'exécution et d'autres intervenants. Le processus d'évaluation peut également faire apparaître que différents participants à une intervention donnée opèrent en fonction d'hypothèses (théories) distinctes quant à la façon dont leur action servira le changement en faveur de la paix. En dégagant les théories sous-jacentes et en ouvrant le débat sur la validité des différentes théories, le processus d'évaluation peut aider les participants à un programme à parvenir à un consensus quant à leur action et aux raisons qui la motivent. Cette clarté permettra éventuellement d'améliorer la conception et l'exécution du programme.

L'évaluation de l'action des donneurs dans les situations de fragilité ou de conflit couvre très souvent plusieurs projets et – dans certains cas – plusieurs donneurs. Cette action fait parfois intervenir de multiples théories ou diverses variantes d'une notion générale, par exemple celle selon laquelle une intensification de l'aide aux services sociaux renforce la stabilité.

Encadré 4.1. Évaluer les réussites et les échecs en matière de construction de la paix

Dans le cadre d'une évaluation, il convient d'établir une distinction entre les différents types de réussite et d'échec et de les analyser de manière à en tirer des enseignements dont il sera tenu compte dans l'élaboration de programmes ultérieurs. Dans ce domaine, il est utile de distinguer un échec de la théorie (hypothèse erronée de la façon dont le changement interviendra) d'un échec de l'exécution. L'échec d'une théorie dénote l'échec d'une activité de prévention du conflit ou de construction de la paix résultant d'une relation causale infondée – autrement dit, les hypothèses sous-jacentes quant aux moyens d'apporter le changement dans ce contexte sont fausses. Une théorie erronée du changement pourrait reposer sur une analyse imparfaite du conflit, ou découler de priorités mal choisies ou d'objectifs contradictoires.

L'échec de l'exécution résulte d'un problème de mise en œuvre de l'activité elle-même (ressources/produits, compétences du personnel, calendrier, localisation, sécurité ou budget) ou des systèmes de gestion. Il s'agira parfois d'évolutions soudaines du conflit qui perturbent ou neutralisent les progrès, en dépit d'une conception par ailleurs satisfaisante de l'activité.

Source : OCDE (1999).

Collecte de données

Comme indiqué en introduction, les évaluateurs et les responsables de l'évaluation découvrent parfois des failles dans les politiques, les stratégies ou les interventions. Il se peut que les objectifs soient imprécis ou n'aient pas été établis, que la théorie du changement ou la logique d'intervention aient été mal formulées (comme décrit plus haut), que certains indicateurs manquent, ou que l'on ne dispose pas de données de suivi ou d'informations de référence. Les conditions de sécurité peuvent avoir des conséquences non négligeables en termes de données, et risquent notamment d'empêcher les évaluateurs de voyager dans certaines régions ou pays et d'accéder aux communautés concernées, aux bénéficiaires du programme et à d'autres informateurs clés.

Les équipes d'évaluation, qui travaillent en coopération avec les organismes responsables, doivent résoudre le problème des données insuffisantes ou manquantes. Elles doivent réfléchir à des moyens de (re)construire les données de référence et autres données manquantes ou de compenser leur absence pendant la procédure d'évaluation, tout en gardant à l'esprit que rien ne peut jamais pleinement remplacer une planification solide. Diverses publications présentent des conseils sur la façon de compenser ou de pallier ces lacunes sans compromettre la qualité de l'évaluation (Banque mondiale, 2005b ; Bamberger et al., 2006 ; OCDE, 1999). L'encadré 4.2 examine un exemple de la façon dont un projet financé par les donateurs en Afghanistan s'efforce d'améliorer la disponibilité et la qualité des données.

Encadré 4.2. **Établir des statistiques de suivi et d'évaluation dans la province de Helmand**

Le programme de suivi et d'évaluation de Helmand (HMEP) est un programme financé conjointement par le Department for International Development du Royaume-Uni (DFID) et l'Équipe provinciale de reconstruction (EPR). Il a pour objectif d'améliorer l'exécution et l'efficacité des programmes de stabilisation et de développement à Helmand (Afghanistan). Le HMEP collecte des données auprès de sources tierces et s'appuie sur une capacité de recherche spécialisée pour établir les cadres de référence et assurer le suivi des indicateurs de progrès au regard du plan défini pour la province. Les données sont présentées sous forme graphique et géospatiale, et stockées dans une base de données interactive accessible à partir d'un site Internet. Le HMEP publie des rapports trimestriels de suivi et d'analyse, et jusqu'à quatre rapports de suivi réactif *ad hoc* par an, conformément aux obligations de l'EPR et du DFID en matière d'information. Le HMEP est par ailleurs conçu pour soutenir le développement d'une capacité programmatique au sein de l'EPR et l'intégration du suivi et de l'évaluation à la planification et à l'exécution du programme.

Source : Coffey International Development, « Helmand Monitoring and Evaluation Programme », <http://uk.coffey.com/our-projects/helmand-monitoring-and-evaluation-programme>.

Une phase décisive de la procédure d'évaluation consiste à recueillir des données permettant de suivre « l'histoire » d'un programme ou d'une politique et leurs effets, notamment leur influence sur les points de vue des participants. Pour recueillir des informations, les évaluateurs font appel aux documents et rapports du programme, à toutes les données de suivi disponibles, à des entretiens avec le personnel du programme, les organisations partenaires, les fonctionnaires locaux, les groupes cible, les participants, les tierces parties (les communautés voisines non ciblées par un programme, par exemple)

et les acteurs internationaux et nationaux qui observent l'intervention à distance (organisations de la société civile, instituts de recherche, donateurs, médias, universitaires, groupes de réflexion).

Les stratégies de collecte de données sont fonction de la conception de l'évaluation, de ses finalités et des sources d'informations disponibles. On citera notamment l'échantillonnage aléatoire et l'échantillonnage raisonné. Des données quantitatives et qualitatives peuvent être recueillies dans le cadre de recensements, d'observations, d'enquêtes auprès des ménages, d'entretiens, de questionnaires, d'études anthropologiques ou ethnographiques, d'ateliers participatifs et de groupes de discussion. Les systèmes statistiques nationaux et les grandes ONG disposent souvent de données démographiques, qui ne sont cependant pas toujours adaptées à la taille ou aux particularités requises de l'échantillon. Par ailleurs, des instituts de sondage privés ont entrepris de se spécialiser dans les enquêtes auprès des ménages et la collecte de données aux normes professionnelles dans les zones en proie à un conflit. L'encadré 4.3 présente des exemples de sources de données et les ouvrages spécialisés en matière d'évaluation fournissent d'autres informations sur les techniques de collecte de données.

Les climats de violence persistante et de fortes tensions posent des problèmes substantiels en termes de disponibilité des données, problèmes qui limitent souvent les travaux des évaluateurs. Il se peut toutefois que davantage de données soient disponibles que ne le laisseraient penser au premier abord la phase de planification et les documents du programme. La rotation élevée du personnel nuit souvent à la mémoire institutionnelle. En outre, les différents intervenants extérieurs et nationaux sont souvent mal coordonnés, et ne partagent pas toujours les informations qu'ils recueillent avec d'autres participants n'ayant peut-être pas connaissance des études, évaluations, données, examens et enquêtes disponibles ou n'y ayant pas accès. De précieuses données de référence, dont peu connaissent l'existence, dorment peut-être quelque part dans des archives. Il vaut parfois la peine de consacrer du temps, en début d'évaluation, au recensement des données déjà collectées, soit dans le cadre de la phase de démarrage de l'évaluation, soit dans celui d'une étude préalable.

Les évaluateurs et autres chercheurs devraient diffuser et utiliser les données rassemblées par d'autres, car trop d'informations sont recueillies qui ne sont pas exploitées par la suite. Il est également recommandé que les commanditaires de l'évaluation et l'équipe chargée de la conduire établissent conjointement des mécanismes d'urgence pour faire face à d'éventuels renversements brutaux de la situation sur le terrain susceptibles d'influer sur la collecte des données nécessaires à l'évaluation. Les responsables de l'évaluation peuvent intégrer au cahier des charges des dispositions en matière de gestion des données.

Pendant un conflit et dans les phases qui lui font immédiatement suite, quand la méfiance est à son comble et que la plupart des parties prenantes ont des intérêts en jeu, la fiabilité (sans parler de la disponibilité) des données et des informations fournies est souvent particulièrement aléatoire. Selon leurs propres intérêts dans le conflit, différents intervenants peuvent avoir des interprétations diverses, voire contradictoires, des impacts positifs ou négatifs d'une intervention et de sa pertinence. Pour garantir la fiabilité des données, les évaluateurs doivent faire appel à plusieurs sources ou types d'informations et à un dosage équilibré de données quantitatives et qualitatives. Ils doivent recouper les données qu'ils utilisent, vérifier la transparence des sources et la validité des données

Encadré 4.3. **Somalie et Afghanistan : exemples de sources d'information**

Somalie : utilisation de plusieurs sources d'information et prise en compte de l'insuffisance des données

Une évaluation conduite en 2011 par DARA (un organisme indépendant spécialisé dans l'évaluation humanitaire) s'est penchée sur l'efficacité de l'intervention humanitaire en Somalie. Elle s'est appuyée sur les sources suivantes :

- les études sur les programmes d'assistance antérieurs en Somalie et les variables contextuelles, 300 pages Internet et publications pertinentes, et divers documents ;
- des entretiens de groupe et individuels semi-structurés avec 489 intervenants, dont 189 (112 au Kenya et 77 en Somalie) ont été conduits avec des individus ayant participé à l'intervention, notamment des représentants des organismes des Nations Unies, de la Croix-Rouge et du Croissant-Rouge, des ONG internationales et nationales, des acteurs publics locaux et des donateurs. Vingt-quatre pour cent des personnes interrogées étaient des femmes. Des entretiens ont été menés avec des groupes de plus de 300 personnes des populations concernées ;
- des observations sur le terrain, au Kenya et en Somalie, dans les camps pour personnes déplacées et pour les réfugiés somaliens ;
- une enquête en ligne conduite auprès du personnel et d'acteurs clés de l'intervention humanitaire (le taux de réponse a été faible, 10 réponses seulement ayant été reçues).

Au total, l'équipe d'évaluation a recueilli 3117 éléments d'information, mais s'est heurtée aux difficultés suivantes : rotation élevée du personnel, temps de travail sur le terrain limité, insécurité, manque de cohérence des données, ventilation insuffisante des données (disponibilité restreinte de données désagrégées selon l'âge, le sexe, les vulnérabilités et les résultats pendant la période examinée). Des problèmes de sécurité ont empêché les évaluateurs internationaux de se rendre dans les régions mentionnées dans le cahier des charges, alors que les évaluateurs nationaux ont été autorisés à se rendre dans certaines d'entre elles.

Nord-est de l'Afghanistan : création de nouvelles données

Plus un État est fragile, moins il y a de données disponibles. Les équipes d'évaluation doivent souvent recueillir leurs propres informations. Dans le contexte d'une évaluation de l'impact de l'aide dans le nord-est de l'Afghanistan, l'équipe a recueilli les données auprès de différentes sources. La plupart ont été collectées dans le cadre de deux enquêtes conduites auprès de 2 000 chefs de famille dans 80 villages. Les évaluateurs en ont également recueilli auprès d'autres sources afin d'établir pour chaque communauté un profil comportant des informations sur l'histoire, la démographie, la composition ethnique, l'organisation politique et sociale et la dotation en ressources de chacune d'elles. Dans d'autres contextes, ces données auraient pu être facilement dégagées de sources existantes, un recensement national par exemple. Dans le contexte afghan, en revanche, les évaluateurs ont dû les recueillir eux-mêmes, ce qui n'a été possible que dans le cadre d'un projet de recherche pluriannuel. Pour rassembler les informations relatives aux principaux événements et changements intéressants ces communautés, un système de suivi et de notification a été établi couvrant 40 villages. Les correspondants locaux ont rempli des rapports semi-structurés quatre fois par an. L'équipe d'évaluation a recueilli d'autres données qualitatives au cours de visites sur le terrain. Enfin, les évaluateurs ont obtenu des données sur les apports d'aide auprès de différents organismes internationaux de développement dans la région. Les données ont ensuite été compilées et utilisées pour procéder à une analyse quantitative et qualitative de l'impact de l'aide sur la stabilité.

Source : Polastro et al. (2011) et Böhnke et al. (2010).

avant de les analyser – en vérifiant les faits auprès des principales parties prenantes et des personnes interrogées, par exemple. Le recours à plusieurs sources de données leur permettra de neutraliser le biais associé à l'utilisation d'un seul type d'information et d'observateurs isolés. Le rapport d'évaluation doit, au chapitre de la description des méthodes, expliquer tous les problèmes rencontrés à cet égard (y compris l'insuffisance des données ou les problèmes d'incompatibilité) de même que leurs répercussions sur la fiabilité et la validité des conclusions de l'évaluation.

Pour éviter d'attiser les tensions entre les groupes (conformément au principe « ne pas nuire »), les décisions concernant les moyens de mobiliser la participation de différents groupes à la collecte de données doivent se fonder sur une compréhension précise des rôles et des intérêts des parties prenantes, et s'appuyer sur les connaissances acquises au sujet de ces dernières dans le cadre de l'analyse du conflit. Lorsque des problèmes d'accès ou de sécurité font obstacle à la collecte de données, d'autres méthodes, notamment des consultations avec des remplaçants informés en mesure de fournir des avis ou des informations représentatives, sont envisageables. Il conviendra éventuellement de prêter attention à la couverture des données – de manière à tenir compte de points de vue exprimés à l'extérieur et à l'intérieur de la capitale par exemple – en fonction des éléments sur lesquels doit porter l'évaluation.

Les évaluations doivent se fonder sur des données précises et pertinentes sur les femmes, les hommes et les relations entre les sexes. Cela garantit la prise en compte questions de genre et de la situation de conflit, met en évidence des disparités entre les sexes et aide à répondre aux questions de l'évaluation. Des données statistiques ventilées par sexe et l'analyse de genre (renseignements sur la situation des femmes et des hommes dans un contexte de conflit particulier) doivent informer l'évaluation. En l'absence de données désagrégées, on peut associer des approches quantitatives et qualitatives pour dégager les éléments nécessaires à la compréhension des relations entre les sexes et différencier les résultats pour les hommes, les femmes, les garçons et les filles. Si l'intégration de ces données est jugée inutile ou irréalisable, les raisons doivent en être clairement exposées dans le rapport d'évaluation. Ainsi, dans le cadre d'une évaluation de l'impact de l'aide sur l'état d'esprit des populations locales dans le nord-est rural de l'Afghanistan (Böhnke et al., 2010), l'équipe d'évaluation a procédé à une enquête. Elle souhaitait y faire participer les femmes et les jeunes mais, au bout du compte, seuls les chefs de famille de sexe masculin ont été interrogés. Cette décision s'est fondée sur une analyse du contexte local, sur des consultations avec des experts régionaux et des spécialistes de la problématique genre, et sur une expérience antérieure ayant fait appel à des chercheuses pour mener une enquête auprès des femmes afghanes, celles-ci ayant estimé qu'en raison des menaces pour leur sécurité et de la coutume exigeant que chaque chercheuse soit accompagnée d'un homme, il était quasiment impossible pour les équipes de travailler efficacement. Les évaluateurs ont décidé que le choix de ne pas intégrer les femmes à l'enquête était justifié dans ce cas parce que les opinions des hommes plus âgés (en qualité de représentants politiques des ménages) étaient celles qui présentaient le plus grand intérêt pour répondre aux questions clés de l'évaluation. Ils ont vérifié leur décision auprès d'équipes locales expérimentées et l'ont testée avant d'établir le cadre de référence initial. Dans son rapport d'évaluation, l'équipe a expliqué sa décision de la façon suivante : « bien que nous estimions que les travaux à conduire justifient ces choix et les limitations qui en résultent, nous sommes bien conscients que la conception de notre étude ne permet pas d'appréhender les tendances pour l'ensemble de l'Afghanistan, pas plus qu'elle

n'appréhende la manière de voir des femmes [...] » (Böhnke et al., 2010). La transparence quant aux méthodes et aux limitations des données est un élément indispensable à une évaluation de qualité.

De nombreuses interventions s'efforcent de consolider la paix et de prévenir les conflits en suscitant une évolution des comportements, des processus de pensée et des relations parmi les populations. Dans ces cas, il peut s'avérer nécessaire de recueillir des données comportementales, de conduire des entretiens, des ateliers ou des discussions de groupe avec les parties intéressées, ou de mener des enquêtes en vue de collecter des données quantitatives. Le même processus de vérification par triangulation des données est requis pour l'évaluation de changements intangibles, par exemple pour apprécier les perceptions au travers d'entretiens.

Il arrive parfois qu'il soit impossible d'interroger les personnes voulues (sources primaires) parce qu'elles sont en prison, décédées ou injoignables. Le rapport d'évaluation doit rendre compte de ces lacunes et expliquer comment il y a remédié. En faisant le point sur les données disponibles au début de la phase de préparation, on évitera des surprises ultérieures de cette nature, qui peuvent compromettre ou retarder le processus d'évaluation.

Critères d'évaluation des interventions

Lorsque l'on évalue les programmes et projets de coopération pour le développement, il est utile de prendre en considération les critères énoncés dans les *Principes du CAD pour l'évaluation de l'aide au développement* et d'autres critères susceptibles de présenter un intérêt. Leur analyse constitue l'essentiel d'un rapport d'évaluation. La présente section montre comment adapter chacun des critères d'évaluation du CAD de l'OCDE aux situations de conflit ou de fragilité. Les questions énoncées ci-dessous sont données à titre de référence et ne constituent nullement une liste exhaustive ou obligatoire. Les critères d'évaluation sont étroitement liés, chacun étant évalué sous un angle légèrement différent de manière à dresser un tableau aussi complet que possible de l'intervention. Considérés dans leur ensemble, ils devraient aider l'équipe d'évaluation à se faire une idée précise de l'activité ou de la politique évaluée et de sa contribution au renforcement de l'État et à la construction de la paix. Cette adaptation des critères se fonde sur les contributions de Paffenholz et Reychler (2007), l'ébauche de lignes directrices (OCDE, 2008b), OCDE et CDA (2007), et Anderson et Olson (2003).

Pertinence

Le critère de pertinence sert à évaluer dans quelle mesure les objectifs et les activités de la (des) intervention(s) répondent aux besoins des bénéficiaires et du processus de construction de la paix – autrement dit, s'ils remédient aux principaux facteurs déterminants du conflit que l'analyse aura révélés. Il associe les effets directs de l'analyse du conflit aux objectifs de l'intervention, bien que la pertinence de l'intervention puisse changer à mesure que la situation évolue. Pour apprécier la pertinence, il conviendra éventuellement d'évaluer dans quelle mesure une intervention s'articule avec les stratégies globales et le cadre politique du pays ou des partenaires extérieurs. Différents groupes ou acteurs du conflit peuvent avoir des points de vue divergents sur la pertinence d'une intervention et ses résultats. La perception des hommes peut également différer de celles des femmes.

Il est indispensable d'évaluer une intervention au regard du conflit pour mesurer sa pertinence. Si le personnel, les responsables et d'autres personnes participant à sa conception et à son exécution ont déjà procédé à une analyse du conflit, il conviendra d'en évaluer la précision et l'utilisation. Un volet majeur de l'évaluation consistera à déterminer si cette analyse s'avère (ou s'est avérée) fondée. Cet exercice permettra d'acquérir des connaissances et de peaufiner les théories quant aux raisons de la violence et aux déterminants majeurs de la dynamique du conflit et du renforcement de l'État à long terme. Si aucun processus d'analyse systématique n'a été suivi auparavant, l'équipe d'évaluation peut soit s'entretenir avec le personnel et les parties prenantes pour comprendre la perception sous-jacente (non formulée) du conflit guidant leurs travaux, soit organiser un exercice plus formel d'analyse du conflit. L'analyse est nécessaire pour définir les liens escomptés entre les résultats du programme et les objectifs de construction de la paix ainsi que leur pertinence par rapport aux principaux facteurs de conflit ou de fragilité.

Les questions permettant de déterminer la pertinence pourraient être les suivantes :

- L'intervention est-elle fondée sur une analyse valide de la situation de conflit ou de fragilité ? A-t-elle été progressivement adaptée pour tenir compte des données actualisées tirées des analyses ?
- Compte tenu de l'analyse du conflit, l'intervention traite-t-elle les bonnes questions dans ce contexte et à ce moment précis ? Semble-t-elle agir sur les causes et les facteurs essentiels du conflit ou de la fragilité ? Ou est-elle axée sur le comportement des principaux groupes à l'origine du conflit ?
- Quel jugement la population locale, les bénéficiaires et les observateurs extérieurs portent-ils sur la pertinence de l'intervention ? Cette perception varie-t-elle selon le sexe ?
- Les buts et objectifs affichés sont-ils pertinents au regard des problèmes qui sont au cœur de la situation de conflit et de fragilité ? Les activités et les stratégies sont-elles adaptées aux objectifs – autrement dit, existe-t-il une cohérence interne entre ce qui est fait dans le cadre du programme et ce qui est recherché comme résultat ? L'intervention s'est-elle progressivement adaptée à l'évolution de la situation ? L'analyse du conflit a-t-elle été repensée ou actualisée pour guider les activités dans un contexte en constante évolution ?

Efficacité

Le critère d'efficacité sert à évaluer si une intervention a atteint, ou a des chances d'atteindre, les objectifs visés par rapport à l'environnement de construction de la paix qui prévaut. L'essentiel pour évaluer l'efficacité – et donc le lien entre les produits, les effets directs et les impacts – consiste à définir dans quelle mesure les résultats envisagés ont été obtenus et à noter les transformations que l'intervention a déclenchées ou auxquelles elle a contribué. En outre, comme la plupart des opérations conduites dans les situations de conflit ou de fragilité ne visent pas explicitement à instaurer une paix durable, il importe d'établir une distinction entre deux types de résultats : d'une part, « l'efficacité du programme », à savoir dans quelle mesure le programme a atteint l'objectif affiché ; d'autre part, si le programme a rempli ses objectifs ou son but, les résultats immédiats ou secondaires associés à la dynamique de construction de la paix et de conflit définie dans l'analyse.

On peut faire appel à une analyse de la théorie du changement pour évaluer le critère d'efficacité. Il s'agit de déterminer si l'intervention s'appuie sur une théorie et une logique solides et si celles-ci se vérifient (ou se sont vérifiées). La perception de l'efficacité est également associée à l'analyse du conflit. Un programme ou une politique peuvent avoir des effets bénéfiques ou fonctionner correctement sans pour autant modifier la dynamique sous-jacente de la situation de conflit ou de fragilité définis par l'analyse du conflit. Il se peut en outre que des facteurs extérieurs, indépendants de l'activité en question, exercent une influence plus prononcée sur la paix et le conflit, auquel cas l'efficacité sera perçue par rapport à ces dynamiques et évolutions plus générales.

L'encadré 4.4 décrit comment l'Allemagne a évalué l'efficacité de son programme de service civil pour la paix au moyen d'un cadre associant une analyse des théories du changement à des études empiriques.

Encadré 4.4. **Évaluation de l'efficacité du Service civil allemand pour la paix**

L'Allemagne a commandité une évaluation indépendante de son programme de service civil pour la paix qui promeut des méthodes non violentes pour remédier aux conflits à travers un éventail de projets d'enseignement, de formation et de services. Les évaluateurs ont établi un cadre d'évaluation complet qui a permis une comparaison systématique des données concernant huit pays. Pour pallier l'insuffisance des données, diverses approches et méthodes d'évaluation ont été combinées. L'équipe d'évaluation a reconstruit les cadres de référence et décrit les théories du changement sous-jacentes au Service civil pour la paix, et a évalué en outre la plausibilité des effets directs, à savoir la probabilité que l'intervention ait abouti, ou aboutisse ultérieurement, aux effets directs escomptés. Pour ce faire, elle a évalué les interventions du Service civil pour la paix à l'aune d'une liste de critères d'efficacité élaborée durant l'étude empirique. Elle a également dégagé des exemples de bonnes pratiques parmi les interventions évaluées, puis défini les facteurs qui avaient concouru à leur efficacité.

Source : Paffenholz (2011).

Les questions permettant de déterminer l'efficacité pourraient être les suivantes :

- L'intervention a-t-elle atteint son objectif déclaré (ou implicite), ou est-elle raisonnablement susceptible de le faire au vu des produits et effets directs ?
- La théorie du changement est-elle fondée sur des hypothèses valides/vérifiées ? Existe-t-il d'autres théories ?
- Dans quelle mesure les donneurs ont-ils recensé et su gérer les risques contextuels ?
- L'intervention progresse-t-elle selon un calendrier raisonnable, ou le fera-t-elle ? Est-il possible d'accélérer le processus ? Faut-il, pour une raison ou une autre, freiner le déroulement des opérations ?
- Quels sont les principaux facteurs qui concourent à la réalisation ou à la non-réalisation des objectifs ?
- L'intervention a-t-elle abouti à des résultats différents pour les femmes, les hommes, les garçons et les filles ?

Selon l'activité évaluée et la portée de l'évaluation, on pourra poser des questions relatives à l'efficacité dans une situation de conflit, par exemple : l'intervention a-t-elle

pour effet de renforcer la résistance des populations à la violence et à la provocation ? Les parties intéressées ont-elles un impact notable sur la situation de conflit ou de fragilité ? L'intervention couvre-t-elle les populations clés/appropriées ou des populations nombreuses ? La question de l'égalité des sexes et les inégalités horizontales pertinentes (ethniques, religieuses, géographiques, etc.) qui favorisent le conflit ont-elles été prises en considération, et avec quels résultats ? L'intervention a-t-elle amélioré la sécurité des populations et leur sentiment de sécurité ? Cela favorise-t-il les méthodes non violentes de résolution des conflits ou de gestion du pouvoir ? S'est-elle traduite par une véritable amélioration des relations entre les groupes en conflit, perceptible dans l'évolution des comportements ?

Impact

Le critère de l'impact porte sur les effets plus généraux d'une intervention. Ceux-ci peuvent être positifs ou négatifs, produits directement ou indirectement, intentionnellement ou non. Dans les situations de fragilité ou de conflit, le critère de l'impact sert à recenser et à évaluer les effets de l'intervention sur les facteurs et les acteurs déterminants du conflit et, s'il y a lieu, sur les processus plus vastes de développement et de renforcement de l'État. L'évaluation doit couvrir à la fois les transformations visées par l'intervention et tous les résultats fortuits (ou imprévus), positifs ou négatifs.

Les modifications des comportements et des mentalités du type que visent de nombreuses interventions de construction de la paix sont souvent difficiles à mesurer et parfois longues à obtenir. En ce sens, il arrive qu'il soit trop tôt pour pouvoir raisonnablement espérer des impacts sensibles sur les facteurs déterminants du conflit (pour les activités visant à réformer les institutions par exemple). Dans ce cas, l'évaluation pourrait porter sur les effets directs et les impacts à court terme, et vérifier la théorie et la logique du programme pour déterminer si les stratégies en vigueur ont des chances raisonnables de contribuer à la paix sur le long terme.

Les méthodes d'évaluation d'impact sont examinées dans d'autres publications portant sur l'évaluation, par exemple dans les orientations commanditées par le Réseau de réseaux pour l'évaluation d'impact – NONIE (Leeuw et Vaessen, 2009). Les méthodes quantitatives rigoureuses associées à ce type d'évaluation et les essais contrôlés aléatoires sont jugés irréalisables dans de nombreuses situations de conflit ou de fragilité (bien que des expériences utiles soient actuellement conduites au niveau des programmes). Il est dans tous les cas particulièrement difficile d'appliquer ces méthodes à des questions générales de paix et de conflit couvrant différentes interventions à l'échelon des pays ou à des évaluations de l'engagement global des donateurs dans un contexte de conflit. Lorsque la causalité ne peut être déterminée de manière fiable au moyen de méthodes rigoureuses, les évaluateurs peuvent donner des explications plausibles à leurs conclusions quant à l'impact de l'intervention, en veillant néanmoins à expliquer leurs insuffisances. L'encadré 4.5 décrit plus en détail les méthodes quantitatives d'évaluation de l'impact.

En cas de conflit d'une violence extrême, l'analyse devra éventuellement s'orienter sur des fonctions de suivi ou d'évaluation « en temps réel » qui privilégient les indicateurs de produits. Ils sont plus faciles à surveiller et moins susceptibles d'un défaut d'attribution que les indicateurs d'impact.

Encadré 4.5. Méthodes quantitatives d'évaluation de l'impact dans des situations de conflit ou de fragilité

Une évaluation d'impact mesure les impacts nets d'une intervention en comparant ses résultats à une situation contrefactuelle – une mesure de ce qui se produirait si l'intervention n'avait pas lieu. Bien que l'on n'ait guère d'expérience relative à l'évaluation d'impact des interventions de construction de la paix, cette méthode est de plus en plus utilisée dans le cadre du développement, y compris dans les situations de conflit ou de fragilité.

Il existe deux méthodes fondamentales pour conduire une comparaison contrefactuelle. La première est l'approche expérimentale, ou évaluation par essai contrôlé aléatoire. Ces évaluations doivent être conçues et engagées avant le début de l'intervention. Les évaluateurs affectent de manière aléatoire les bénéficiaires potentiels à des groupes de « traitement » et de « contrôle ». La comparaison des impacts entre les deux groupes révèle l'impact net de l'intervention. Cette approche peut être utilisée pour piloter une intervention en vue d'éclairer la conception du programme. La seconde couvre les approches « quasi expérimentales », en vertu desquelles on utilise des techniques statistiques pour construire un groupe de contrôle pouvant servir de groupe contrefactuel. Ces techniques couvrent toute une gamme de méthodes d'appariement, de modèles de variables instrumentales, de modèles de régressions par discontinuité, et de méthodes de panel. Elles estiment l'impact d'une intervention en veillant, dans la mesure où les données le permettent, à ce que les bénéficiaires et non-bénéficiaires soient semblables sur tous les plans à l'exception du traitement attribué. L'une d'elle est la technique de l'appariement selon le score de propension, qui estime l'impact net d'une intervention en appariant les bénéficiaires aux non-bénéficiaires qui leur sont similaires sur tous les autres plans.

Ces méthodes d'évaluation des interventions de construction de la paix présentent à l'évidence des défauts, notamment l'absence fréquente de données de référence, la volonté de « traiter » tout le monde (l'impératif humanitaire), les dangers politiques d'un traitement aléatoire, et la nature qualitative de nombreuses variables pertinentes qui sont en conséquence moins compatibles avec les méthodes statistiques. Les évaluations d'impact sont néanmoins souvent réalisables à l'échelle du projet ou du programme et peuvent fournir des informations utiles sur ce qui fonctionne.

Source : Marie Gaarder (Département d'évaluation de la Norad) et Annette Brown (3iE).

Les questions permettant de déterminer les impacts pourraient être les suivantes :

- Quels sont les effets primaires et secondaires, directs ou indirects, positifs ou négatifs, visés ou imprévus, immédiats ou à long terme, à court terme ou durables, de l'activité ou de la politique en question ? Exerce-t-elle une influence notable sur les facteurs essentiels du conflit ou de la paix ?
- D'après l'analyse du conflit, sur quels facteurs déterminants du conflit et de la fragilité l'activité ou la politique a-t-elle influé, et comment ? Quelles modifications dans les mentalités, les comportements, les relations ou les pratiques peuvent être établies (en nombre de personnes et/ou selon une classification en fonction des critères retenus comme le sexe) ? Des effets négatifs secondaires sont-ils observables ?
- De quelle manière la situation a-t-elle évolué au fil du temps et quelle a été, le cas échéant, la contribution de l'intervention à ces évolutions ?

- Quels ont été les effets des interventions sur les indicateurs spécifiques du bien-être visés, comme l'état de santé ou le niveau de pauvreté ? Quelles sont les retombées sur les trajectoires de développement à long terme ?
- L'intervention a-t-elle influencé l'action publique ? En quoi ces politiques sont-elles associées au conflit ?

Durabilité

On entend par durabilité la poursuite des effets positifs une fois l'intervention terminée. Dans une situation de conflit ou de fragilité, la durabilité recouvre la probabilité que les avantages à long terme et la résilience au risque perdurent, de même que les effets bénéfiques sur l'économie, les institutions, la gestion des ressources humaines, etc. Comme dans d'autres domaines, la durabilité suppose également « l'appropriation » des processus de paix et de développement. L'expérience et les études portant sur la paix démontrent que les processus de construction de la paix sont des procédures de longue haleine, qui appellent un engagement durable susceptible de connaître des revers (OCDE et CDA, 2007). Dans les régions en proie à un conflit, cet engagement suppose d'agir sur ceux qui ont intérêt à perpétuer le conflit (parfois dénommés les « saboteurs », ou les groupes « difficilement accessibles »).

Les questions portant sur la durabilité pourraient être les suivantes :

- Quelles sont les mesures qui ont été prises ou qui sont envisagées pour créer des processus, des structures, des normes et des institutions durables à l'appui de la paix ? Dans quelle mesure les opérations visant à renforcer l'appropriation et la participation ont-elles porté à la fois sur les hommes et les femmes ?
- Les nouvelles institutions destinées à lutter contre le conflit et la fragilité perdureront-elles ? Sont-elles utilisées ? Par qui ? L'intervention favorise-t-elle un élan en faveur de la paix en encourageant les participants et les communautés à mettre sur pied leurs propres projets ?
- Une stratégie concrète de « transfert » ou de retrait a-t-elle été définie avec les partenaires ou intervenants locaux pour leur permettre d'élaborer ou de poursuivre leurs propres projets de construction de la paix ?

Selon l'activité évaluée et l'ampleur de l'analyse, on pourra poser des questions spécifiquement associées au conflit et portant sur l'efficacité, par exemple :

- L'intervention aboutit-elle à la création ou à la réforme d'institutions ou de mécanismes politiques capables de régler raisonnablement les griefs et les injustices ?
- Ceux qui tirent profit de la violence ou de l'instabilité et ont tout intérêt à ce qu'elles se perpétuent, ou ceux qui s'opposent à la dynamique de paix (les « saboteurs ») ont-ils été convenablement pris en charge ?
- Les améliorations apportées aux relations entre les groupes persisteront-elles face aux nouveaux enjeux et aux nouveaux risques ?
- Les parties à un accord négocié l'honoreront-elles et l'appliqueront-elles ?

Efficience

Le critère d'efficience sert à évaluer dans quelle mesure les ressources économiques (fonds, expertise, temps, etc.) sont convertis en résultats. Dans une situation de conflit, les coûts associés aux travaux de prévention seront souvent comparés aux coûts estimés

d'une guerre ou d'une vague d'affrontements violents. Or, les conflits évités sont invisibles. Il n'existe donc malheureusement pas de situation contrefactuelle pour comparer l'efficacité de la prévention par rapport à un conflit, même si l'on peut exploiter des données historiques pour procéder à une estimation raisonnable des coûts.

Outre la comparaison des coûts d'une action en faveur de la paix et de ceux d'une guerre, l'évaluation pourrait examiner les priorités au regard des principaux facteurs de conflit. Elle pourrait poser la question suivante : « cette méthode particulière de lutte contre la violence est-elle l'option la plus efficace ? » Dans un contexte où les ressources consacrées au développement sont rares, les évaluations devraient indiquer si une activité donnée est « rentable » par rapport à d'autres mesures ou à une non-intervention. L'encadré 4.6 examine les dépenses de sécurité associées à des opérations dans une zone de conflit, en l'espèce l'Afghanistan.

Encadré 4.6. Dépenses de sécurité

Une évaluation conjointe de l'aide humanitaire et de l'assistance à la reconstruction apportées en Afghanistan entre 2001 et 2005 a établi à environ 20 % le pourcentage des frais généraux imputables à la sécurité – autrement dit, les coûts de fonctionnement en Afghanistan ont été supérieurs de 20 % à ceux de projets similaires dans des environnements plus stables. L'évaluation a indiqué qu'en raison de ces coûts inattendus (ou sous-estimés), les programmes conduits en Afghanistan étaient considérablement plus onéreux que des programmes analogues conduits ailleurs. D'autres analyses et d'autres données seraient cependant nécessaires pour déterminer s'ils peuvent néanmoins être considérés comme efficaces.

Source : Chr. Michelsen Institute (2005).

Lorsqu'ils examinent les interventions humanitaires ou de développement dans une région en proie à un conflit, les évaluateurs doivent en déterminer l'efficacité au regard d'autres formes d'appui au développement et à la construction de la paix dans ce contexte ou dans des contextes similaires. Les questions relatives à l'efficacité pourraient être les suivantes :

- Les résultats de l'intervention sont-ils obtenus de façon efficace par rapport au scénario contrefactuel ?
- Dans quelle mesure les ressources (humaines, financières, administratives) sont-elles correctement utilisées pour obtenir des résultats ?
- Existe-t-il de meilleurs moyens (plus efficaces) d'atteindre les objectifs ?
- Quelles mesures ont été prises pour garantir l'efficacité de l'intervention par rapport à son coût ? L'intervention s'est-elle substituée à des projets locaux ou est-elle venue les compléter ?

Autres critères : cohérence et coordination

Une bonne part des études consacrées à la construction de la paix, au renforcement de l'État et à l'efficacité de l'aide – et, de plus en plus, les politiques gouvernementales elles-mêmes – indiquent qu'il est nécessaire d'améliorer la cohérence et la coordination des approches. Le financement alloué à un projet particulier de prévention de conflit ou de

construction de la paix peut-être occulté ou contrecarré ou, à l'inverse, favorisé et soutenu par d'autres interventions engagées par le (les) même(s) gouvernement(s). Les données concrètes concernant l'utilisation et l'utilité de ces approches demeurent cependant insuffisantes. Le cahier des charges de l'évaluation pourrait donc demander aux évaluateurs d'examiner ce contexte plus vaste et les charger de définir dans quelle mesure les activités sont compatibles avec d'autres politiques ou mesures prises par d'autres organismes publics. Les évaluations peuvent également analyser le degré de coordination entre les donateurs, l'administration publique et d'autres intervenants.

Les évaluations peuvent mesurer les coûts et avantages d'un investissement destiné à améliorer la coordination et la cohérence et toutes les conséquences imprévues en découlant. Il convient de ne pas supposer automatiquement que la coordination avec d'autres intervenants contribue à améliorer les résultats. En effet, les approches faisant appel à une plus grande coordination des donateurs peuvent s'accompagner des risques suivants : des structures de gestion lourdes qui réduisent la flexibilité et augmentent les frais généraux ; de fortes pressions sur les parties prenantes nationales résultant d'une position commune des intervenants extérieurs (l'effet « front commun ») ; et des influences inappropriées sur la neutralité, l'impartialité et l'indépendance des acteurs de l'humanitaire et du développement à travers une association avec des acteurs militaires (rétrécissement de l'espace humanitaire). Les évaluateurs doivent soigneusement examiner la réalité des mécanismes de coordination au regard de l'analyse du conflit et d'autres critères (en particulier la pertinence, l'efficacité et l'efficience). Bennett et al. (2010) ont constaté au Soudan du Sud que, dans certains cas, les projets financés par un seul donneur étaient plus efficaces que des mécanismes communs coordonnés.

Pour évaluer la cohérence et la coordination, on pourra poser les questions suivantes :

- La cohérence et la coordination ont-elles été intégrées aux moyens déployés et aux effets directs obtenus (autrement dit, un budget leur a-t-il été attribué et figurent-elles explicitement dans la liste des produits) ?
- Quels ont été les rôles des autres intervenants et les relations avec eux ?
- Quels volets de l'intervention ont fait l'objet d'une bonne coordination ou vu leur cohérence renforcée ?
 - ❖ Analyse conjointe ou compréhension commune du conflit ou de la fragilité dans le contexte de l'intervention ?
 - ❖ Définition commune des priorités de financement ou d'intervention ?
 - ❖ Suppression ou réduction des aspects redondants des programmes ?
 - ❖ Répartition efficace des tâches selon les secteurs, les dossiers ou les problèmes ?
 - ❖ Évaluation commune des programmes à l'échelon du pays ou du secteur ?
 - ❖ Échanges fructueux d'expérience et mise en évidence d'enseignements ?
- La coordination et la cohérence ont-elles amélioré l'efficacité, l'efficience ou les impacts ?
- Combien de temps et quelles ressources ont été consacrés à la coordination ? Cette démarche a-t-elle été efficiente (rapport coût-avantages) et adaptée ? A-t-elle réduit les coûts de transaction ?
- Comment la coordination a-t-elle été établie ? Quels ont été les principaux obstacles et les principales difficultés ? Cette opération peut-elle être reconduite dans d'autres situations (quels facteurs contextuels ont influé sur la coordination et les résultats des efforts de coordination) ?

Dégager des conclusions et formuler des recommandations

L'objectif ultime de toute évaluation est de répondre aux questions clés qui sont posées et de présenter utilement les résultats. Le rapport d'évaluation devrait montrer de façon claire et crédible le lien entre les constats, les conclusions et les recommandations. C'est sur l'analyse des données recueillies que reposeront les conclusions. Ces dernières devraient être présentées aux lecteurs sous la forme d'une démonstration clairement argumentée. Toute recommandation devrait être fondée, pertinente, destinée aux utilisateurs visés, et opérationnelle (autrement dit, que ceux qui reçoivent les informations soient en mesure de leur donner suite) et concourir à améliorer l'activité évaluée et les activités futures. Les conclusions fondées sur des données factuelles peuvent constituer une base utile à des échanges entre les différents intervenants dans les domaines de la prévention des conflits et de la construction de la paix.

Avec le temps, ces conclusions permettront de mieux comprendre le fonctionnement des processus de construction de la paix et de renforcement de l'État, et d'éclairer ainsi les stratégies d'intervention dans une situation de conflit. En particulier, les évaluations peuvent enrichir les connaissances en faisant apparaître l'exactitude (ou l'inexactitude) et l'utilisation des analyses du conflit, ou en montrant si les théories du changement fonctionnent ou pas, et pourquoi. L'encadré 4.7 donne un exemple de la façon dont une évaluation a permis d'élargir les connaissances grâce à des observations éclairant les théories relatives aux relations entre l'aide et les perceptions qu'en ont les populations.

Encadré 4.7. **L'aide peut-elle convaincre des communautés locales ? Vérification de la théorie en Afghanistan**

Dans le cadre d'une évaluation de l'impact de l'aide au développement sur la stabilité dans les communautés rurales du nord-est de l'Afghanistan, l'équipe d'évaluation a examiné deux théories soutenues par de nombreux acteurs du développement travaillant dans les zones de conflit. D'abord, l'hypothèse selon laquelle la stabilité dans une situation de conflit ou d'après-conflit dépend de la légitimité et de l'utilité que la population attribue à l'État et à ses partenaires internationaux, et donc de sa coopération avec eux. Ensuite, l'hypothèse selon laquelle l'aide peut influencer le sentiment des populations locales à l'égard de l'État et des acteurs internationaux. L'équipe d'évaluation a recueilli des données sur les avis de la population rurale et les a mis en correspondance avec des données sur la prestation de l'aide aux communautés. Elle a constaté que l'aide avait peu d'impact sur le sentiment de la population à l'égard des intervenants internationaux, mais qu'elle améliore en revanche leur opinion de l'État.

Source : Böhnke et al. (2010).

Selon le type d'évaluation, il est possible de formuler des conclusions et des recommandations dans un cadre participatif. Les évaluateurs pourraient ainsi présenter des constats liminaires à un groupe de parties prenantes clés avec lesquelles ils pourraient alors collaborer en vue de dégager des conclusions utiles. En revanche, si l'évaluation porte plutôt sur la redevabilité, les évaluateurs seront probablement plus enclins à adopter une approche non participative. L'encadré 4.8 présente diverses recommandations d'une évaluation des interventions dans le domaine de la construction de la paix au Soudan du Sud.

Les évaluateurs découvriront parfois des divergences d'opinion fondamentales non seulement en ce qui concerne les interventions, mais aussi l'intérêt de leurs effets directs et de leurs impacts – surtout parce que les perceptions des individus et des groupes sont fortement déterminées par le conflit et par le rôle qu'ils y ont eux-mêmes joué. Au bout du compte, l'évaluation est un exercice fondé sur des valeurs. Une intervention que certains groupes jugeront réussie sera considérée inutile ou dommageable par d'autres.

Encadré 4.8. **Formulation de recommandations à l'intention des donateurs et du gouvernement du Soudan du Sud**

L'équipe d'évaluation chargée d'évaluer l'aide des donateurs aux opérations de prévention du conflit et de construction de la paix au Soudan du Sud (qui allait bientôt devenir l'État indépendant du Soudan du Sud) entre 2005 et 2010 (Bennett et al., 2010) s'est appuyée sur les conclusions de cet exercice pour formuler des recommandations spécifiques et ciblées à l'intention des donateurs et des autorités nationales. Ayant abouti à la conclusion que les donateurs n'adaptèrent pas leur intervention en fonction d'analyses pointues et nuancées des relations de pouvoir, des causes de la vulnérabilité, des éléments déclencheurs du conflit et des indicateurs de résilience (analyse du conflit), l'équipe d'évaluation a recommandé aux donateurs et à la communauté élargie de l'aide de :

- s'assurer que les programmes, nouveaux ou révisés, sont toujours précédés d'une analyse du conflit établissant le lien entre les dynamiques plus larges et celles spécifiques à la région où se déroule l'intervention ;
- planifier les interventions, en assurer le suivi et les évaluer en fonction des éléments critiques recensés ;
- évaluer la réactivité des interventions aux facteurs de conflit.

L'équipe a défini les facteurs de conflit à traiter en priorité et la façon éventuelle de procéder. Elle a proposé ce qui suit au gouvernement du Soudan du Sud (et aux donateurs qui lui apportent leur soutien) :

- allouer davantage de ressources à la création et au maintien de programmes ayant trait aux moyens de subsistance des jeunes hommes qui, à l'heure actuelle, se laissent trop facilement entraîner dans des activités criminelles ;
- donner les moyens aux autorités traditionnelles (les chefs) de traiter les causes profondes de conflits (y compris les litiges portant sur la terre ou la dot de la mariée) dans leurs tribunaux coutumiers, en proposant des programmes de renforcement des capacités destinés à ces tribunaux ;
- établir des mécanismes de surveillance efficaces afin de contrôler les organismes de sécurité.

Source : Bennett et al. (2010).

Rapports

Des rapports sont établis durant tout le processus d'évaluation (y compris les phases de planification et de démarrage). En règle générale, trois rapports principaux sont attendus, mais l'instance commanditaire de l'évaluation peut définir des obligations particulières en la matière, qui doivent être spécifiées dans le cahier des charges.

Rapport de démarrage

Après avoir conduit l'analyse du conflit et recueilli les informations liminaires, les évaluateurs rédigent un rapport de démarrage décrivant la façon dont l'équipe entend conduire l'évaluation et répondre aux questions énoncées dans le cahier des charges. Le rapport présente les risques et les problèmes, les méthodes qui seront utilisées, les outils de collecte de données, les indicateurs (le cas échéant), les modalités de traitement des questions, les théories du changement (implicites ou explicites), les instruments de sélection des échantillons, les études de cas (si elles n'ont pas été choisies avant la désignation de l'équipe), la structure du rapport et un plan de travail pour les travaux restant à effectuer. Les parties prenantes formulent généralement des observations au sujet du rapport de démarrage, souvent dans le cadre d'un groupe de référence (voir plus haut dans le chapitre 4 pour une analyse détaillée de la phase de démarrage).

Rapport provisoire

Une version préliminaire du rapport d'évaluation fait souvent l'objet d'une vaste diffusion aux fins d'observation, ce qui donne aux parties prenantes l'occasion de formuler des commentaires sur l'évaluation. Le calendrier doit prévoir un délai suffisant pour cet exercice. Certains organismes exigent parfois un rapport de « pré-mission » avant que les équipes ne se rendent sur le terrain (entre le rapport de démarrage et le rapport provisoire) pour encourager la redevabilité, ou d'autres types de rapports pendant les missions sur le terrain.

Rapport final

Bien qu'une évaluation puisse aboutir à de nombreux produits différents, un rapport écrit est presque toujours établi. Les rapports et présentations devront être traduits dans la(les) langue(s) locale(s) concernée(s) pour faciliter les échanges avec l'ensemble des parties prenantes. Le rapport final est adressé à tous les participants. Les publics cibles de la diffusion doivent être définis dès le début du processus. De nombreux organismes publient désormais également le rapport final sur Internet, soit en partie, soit dans sa totalité. Dans tous les cas, la confidentialité et la sécurité des participants à l'évaluation doivent être soigneusement protégées.

Réponses de la direction et mesures de suivi

Les recommandations et conclusions de l'évaluation devraient être systématiquement prises en compte. Les populations ou institutions visées par chaque recommandation (généralement les organismes donateurs, les organismes d'exécution ou les administrations nationales) devront y faire suite et prendre les mesures correspondantes. Le cahier des charges devrait prévoir un processus de notification et de réponse à l'évaluation. Souvent, une réponse et un suivi formels de la part des responsables du programme seront exigés. Il appartient à l'organisme commanditaire de rédiger la réponse de la direction, de diffuser les conclusions et les enseignements dégagés de l'évaluation, et d'engager le processus d'apprentissage. La réponse de la direction et les mesures de suivi revêtiront une forme variable selon les institutions et, parfois, selon le type d'évaluation. La réponse comportera souvent une évaluation de la qualité et des limites de l'évaluation.

Dans le cas d'une évaluation conjointe, une réponse commune ou plusieurs réponses particulières devront être rédigées, selon les exigences des institutions concernées.

L'encadré 4.9 présente une réponse officielle publiée par l'Agence canadienne de développement international (ACDI) en réponse aux conclusions d'une évaluation des interventions dans ce qui était alors le Soudan du Sud.

Encadré 4.9. **Le Canada répond à l'évaluation de l'intervention pour l'appui de la construction de la paix au Soudan du Sud**

L'évaluation sur les interventions au Soudan du Sud (*Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan*, Bennett et al., 2010) a formulé des recommandations à l'intention du gouvernement soudanais et de ses partenaires internationaux, dont le Canada. Le Canada appliquant une approche pangouvernementale dans ce pays, tous les départements concernés ont analysé les conclusions du rapport. Par ailleurs, le programme relatif au Soudan de l'Agence canadienne de développement international (ACDI) a donné suite à l'évaluation par la publication d'une réponse officielle de l'équipe de gestion. Celle-ci établit que l'ACDI a considéré que le rapport d'évaluation était un document analytique utile. Elle a également souligné que l'évaluation a constitué une occasion unique d'apprentissage puisqu'il s'agissait de la première évaluation de programme associant plusieurs donneurs à laquelle l'ACDI avait participé.

Les responsables ont décrit les mesures qu'ils avaient prises ou prévoyaient de prendre en ce qui concernait des recommandations précises. À titre d'exemple, s'agissant de la recommandation de procéder à une meilleure analyse du conflit, ils ont précisé qu'ils « [reconnaissaient] l'importance d'analyser le conflit tout au long de la durée de vie d'un projet » et décrit les mesures engagées pour promouvoir la prise en compte du conflit dans les programmes, dont une cartographie du conflit appelée à orienter la conception et la mise en œuvre de nouveaux projets concernant les enfants et les jeunes ainsi que la sécurité alimentaire. Plus particulièrement, dans le cadre de projets auxquels ont participé le Fonds des Nations Unies pour l'enfance et l'Organisation des Nations Unies pour l'alimentation et l'agriculture, en collaboration avec l'Organisation des Nations Unies pour le développement industriel, un atelier de deux jours avait été prévu pour la réalisation d'un exercice collectif de cartographie destiné à sélectionner les régions où pourraient être réalisés les projets prévus, ainsi qu'à trouver des façons de s'attaquer aux causes profondes du conflit. Les facteurs liés aux besoins en matière de rétablissement rapide ainsi que les possibilités de création de synergies entre les programmes de sécurité alimentaire et de développement des jeunes ont été pris en considération.

Source : Agence canadienne de développement international (2011).

Diffusion des constats

Les projets de suivi et de diffusion des enseignements dégagés devraient être mis en œuvre, comme convenu durant le processus de planification de l'évaluation décrit au chapitre 3. Il convient d'utiliser des moyens appropriés pour communiquer les résultats au(x) groupe(s) cible(s), comme les responsables, le personnel et les décideurs. Les responsables de l'évaluation recourent de plus en plus à de brefs résumés, à des traductions en langue locale, à des synthèses, à des clips vidéo et à d'autres instruments de communication adaptés pour communiquer aux différents publics les constats qui les concernent.

Il est parfois difficile de diffuser les conclusions d'une évaluation lorsque les résultats sont perçus comme négatifs ou lorsqu'ils mettent en question des stratégies ou des

approches auxquelles les experts sont très attachés. Les intervenants refusent parfois de voir l'efficacité de leur méthode remise en cause. On peut améliorer la réceptivité en faisant valoir la dimension d'apprentissage d'une évaluation, et en mobilisant les parties prenantes dès les premières phases.

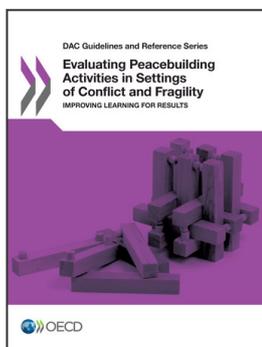
Intégration à la programmation et mise en œuvre d'un processus d'apprentissage

Lorsqu'une évaluation est achevée, ses conclusions devraient être intégrées aux premières phases de la planification et de la conception des programmes et contribuer à résoudre des difficultés en apportant des preuves de la validité (ou pas) des théories du changement ainsi que des données à des fins de comparaison et de référence. Les évaluations conduites pendant l'exécution d'un programme ou d'une stratégie peuvent servir à ajuster ou à recentrer ces derniers.

Outre l'utilisation immédiate des observations de l'évaluation (par ses commanditaires par exemple), les données, les enseignements et les conclusions qu'elle présente pourront éventuellement alimenter d'autres forums et activités de recherche. Par exemple, l'évaluation d'un programme travaillant avec des femmes médiatrices pour la paix livrera peut-être des informations sur la façon dont les femmes sont perçues dans les différentes communautés en conflit qui pourront s'avérer utiles pour travailler dans le contexte du conflit. Elle pourra aussi livrer des enseignements pour la mise en œuvre efficace de programmes de médiation, ainsi que des facteurs contextuels de réussite éventuellement utiles aux personnes travaillant dans le domaine de la médiation dans des environnements très différents.

Les évaluateurs devraient, en particulier, faire preuve de transparence quant à la validité extérieure des constats de l'évaluation, c'est-à-dire expliquer dans quelle mesure ils sont applicables dans d'autres contextes. L'utilisation erronée ou inadéquate de constats peut avoir des conséquences très négatives. C'est pourquoi il importe que les évaluateurs décrivent explicitement les points forts et les points faibles des données et méthodes utilisées et la façon dont ils influent sur les constats, les conclusions et les recommandations.

Une fois les processus d'évaluation et d'apprentissage terminés, les décideurs, les responsables et le personnel devraient être mieux outillés pour comprendre et améliorer les stratégies, les effets directs et les impacts, et apporter ainsi une contribution plus durable à la paix.



Extrait de :

Evaluating Peacebuilding Activities in Settings of Conflict and Fragility

Improving Learning for Results

Accéder à cette publication :

<https://doi.org/10.1787/9789264106802-en>

Merci de citer ce chapitre comme suit :

OCDE (2013), « Conduire une évaluation dans les situations de conflit ou de fragilité », dans *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility : Improving Learning for Results*, Éditions OCDE, Paris.

DOI: <https://doi.org/10.1787/9789264106833-8-fr>

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les arguments exprimés ici ne reflètent pas nécessairement les vues officielles des pays membres de l'OCDE.

Ce document et toute carte qu'il peut comprendre sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Vous êtes autorisés à copier, télécharger ou imprimer du contenu OCDE pour votre utilisation personnelle. Vous pouvez inclure des extraits des publications, des bases de données et produits multimédia de l'OCDE dans vos documents, présentations, blogs, sites Internet et matériel d'enseignement, sous réserve de faire mention de la source OCDE et du copyright. Les demandes pour usage public ou commercial ou de traduction devront être adressées à rights@oecd.org. Les demandes d'autorisation de photocopier une partie de ce contenu à des fins publiques ou commerciales peuvent être obtenues auprès du Copyright Clearance Center (CCC) info@copyright.com ou du Centre français d'exploitation du droit de copie (CFC) contact@cfcopies.com.