

3

Utiliser les critères d'évaluation dans la pratique

Le présent chapitre traite des aspects pratiques liés à la manière dont les critères devraient être utilisés dans le cadre de la conception et de la mise en œuvre des évaluations. Il examine les deux principes fondamentaux qui devraient guider l'application des critères. Après avoir analysé différentes approches susceptibles de favoriser une application réfléchie des critères, il s'intéresse à la façon dont les critères peuvent être appliqués dans différents contextes institutionnels et avec des priorités stratégiques, des méthodes de travail et des cultures différentes. Il décrit sommairement comment les critères peuvent inciter les évaluateurs à examiner les différences en termes d'expérience et d'impact en appliquant un prisme sexospécifique. Enfin, il examine comment les critères peuvent aider les évaluateurs et les responsables d'évaluation à travailler d'une manière qui contribue à la réalisation des Objectifs de développement durable (ODD) et, plus globalement, au Programme de développement durable à l'horizon 2030. Divers exemples concrets ponctuent le chapitre, qui montrent comment les critères ont été appliqués dans le cadre de diverses évaluations.

Pour aider les évaluateurs et les personnes chargées de la conception ou de la gestion des interventions à mettre au point des évaluations utiles et adaptées aux différents contextes et parties prenantes, les deux principes suivants ont été élaborés pour guider l'utilisation des critères. Pour éviter que les critères ne soient appliqués mécaniquement – aux dépens de la réflexion critique, de la créativité et de l'appropriation par les participants –, ces principes devraient être pris en compte dès lors que les critères sont utilisés (OCDE, 2019^[1]) :

- **Principe n°1** : Les critères devraient être appliqués de façon réfléchie pour renforcer la qualité et l'utilité des évaluations.
- **Principe n°2** : L'utilisation des critères dépend de la finalité de l'évaluation.

La section qui suit développe ces deux principes et présente d'autres notions clés liées à l'utilisation des critères, notamment la manière de les adapter à des contextes particuliers, la manière de les examiner à différents moments dans le temps et la manière dont ils sont reliés les uns aux autres.

Une application réfléchie des critères

Le Principe n°1 souligne que les critères devraient être appliqués de façon réfléchie. Dans la pratique, cela signifie mener une réflexion critique sur les critères qui sont le plus utiles pour contribuer à une évaluation utile et de qualité, présentant un intérêt pour les utilisateurs auxquels elle est destinée. L'Encadré 3.1 illustre la manière dont les critères peuvent être appliqués de façon réfléchie lors de l'évaluation d'une intervention.

La prise en compte des six aspects suivants et des questions qui leur sont associées aidera les évaluateurs à utiliser les critères de façon réfléchie :

- **Contexte** : Quel est le contexte de l'intervention et comment les critères peuvent-ils être appréhendés dans le contexte de l'évaluation, de l'intervention et des parties prenantes concernées ?
- **Finalité** : Quel est le but que l'évaluation tente d'atteindre et quelles questions sont les plus utiles pour poursuivre et atteindre ce but ?
- **Rôles et dynamiques de pouvoir** : Qui sont les parties prenantes, quels sont leurs besoins et intérêts respectifs ? Quelles dynamiques de pouvoir sont à l'œuvre ? Qui doit être associé à la prise de décisions sur les critères à appliquer et comment appréhender ces critères dans le contexte spécifique de l'évaluation et de l'intervention ? Peuvent ici être posées des questions sur l'appropriation et sur les personnes qui décident de l'objet et des priorités de l'évaluation.
- **Intervention (sujet évalué)** : Quel type d'intervention est évalué (un projet, une politique, une stratégie, un secteur) ? Quelle est sa portée et sa nature ? Dans quelle mesure ses résultats attendus sont-ils directs ou indirects ? Quelle théorie des systèmes complexes entre en jeu ?
- **Évaluabilité** : Existe-t-il des contraintes en termes d'accès, de ressources et de données (y compris les données ventilées) qui ont une incidence sur l'évaluation, et en quoi ces contraintes influent-elles sur les critères ?
- **Temporalité** : À quel stade du cycle de vie de l'intervention l'évaluation sera-t-elle effectuée ? Le contexte dans lequel s'inscrit l'intervention a-t-il évolué avec le temps et, dans l'affirmative, de quelle manière ? Ces évolutions devraient-elles être prises en compte pendant l'évaluation ? Le moment choisi pour l'évaluation influera sur l'utilisation des critères ainsi que sur les sources de données.

Encadré 3.1. Utilisation réfléchie des critères : évaluation de la dotation de la Norvège en faveur de la société civile dans les pays en développement

Cette évaluation visait à examiner la dotation de la Norvège en faveur de la société civile, qui a pour objet de renforcer la société civile dans les pays en développement et, partant, de contribuer à une société civile plus forte, dotée de la capacité et des moyens nécessaires pour promouvoir la démocratisation, améliorer les droits humains et réduire la pauvreté.

Dans le cadre de cette évaluation, les critères ont été appliqués de façon « réfléchie » et souple, et interprétés d'une manière particulière tenant compte de la nature de l'intervention (partenariats avec des organisations de la société civile et renforcement de leurs capacités) ainsi que des priorités norvégiennes et du contexte. L'évaluation s'appuyait sur les critères suivants, qui avaient été adaptés à certains égards :

- impact – défini au regard des principales réalisations, par exemple la démocratie ou les possibilités de revenus entraînant une réduction de la pauvreté
- pertinence – aucun changement
- efficacité – définie au regard des objectifs spécifiques du partenariat (c'est-à-dire prestation de services, sensibilisation et renforcement des capacités). Englobait certaines réalisations à moyen terme, comme la création d'un espace dévolu à la société civile
- viabilité/durabilité – aucun changement.

Un autre critère, celui de la « valeur ajoutée », était également utilisé, défini comme la compétence professionnelle, la compétence organisationnelle et financière et la compétence relationnelle (réseaux). Ce critère se rapportait à des aspects qui, dans d'autres contextes, pourraient relever des critères d'efficacité et d'efficience.

Source : Helle et al. (2018^[2]), *From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations*, https://www.norad.no/globalassets/filer-2017/evaluating/1.18-from-donor-to-partners/1.18-from-donors-to-partners_main-report.pdf

Adapter les critères à la finalité de l'évaluation

Lorsqu'il s'agit de décider de la manière dont les critères devraient être utilisés, le plus important est de les relier à la finalité de l'évaluation et à son contexte, puis de les articuler, ainsi que les questions d'évaluation, autour de cette finalité. L'Encadré 3.2 fournit deux exemples de la manière dont la finalité d'une évaluation peut être définie.

Les critères n'ont pas vocation à être appliqués de manière uniformisée ni figée à chaque intervention, ni ne sont censés servir de liste de vérification. En réalité, ils devraient être interprétés ou appréhendés avec circonspection, au regard de l'intervention évaluée. Une telle démarche favorise la souplesse et l'adaptation des critères à chaque évaluation. Il faudrait clairement indiquer sur quelles notions sous-jacentes aux critères l'évaluation s'appuiera en particulier, et pourquoi.

La finalité de l'évaluation devrait être soigneusement et clairement définie. Les parties prenantes qui participent à l'évaluation devraient être associées à ce stade afin de s'assurer qu'elles comprennent le but de l'évaluation et l'utilisation qui en sera faite.

Parmi les principales questions à examiner pour déterminer la finalité de l'évaluation figurent les suivantes :

1. Quel est le besoin d'évaluation, qui est le public cible et quel usage fera-t-il des constatations de l'évaluation ?
2. Quelles actions sont réalisables compte tenu des caractéristiques et du contexte de l'intervention ?
3. Quel degré de certitude est nécessaire pour pouvoir répondre aux questions clés ?
4. À quel moment les informations sont-elles nécessaires ?
5. Que sait-on d'ores et déjà de l'intervention et de ses résultats ? Qui est détenteur de ces informations et comment sont-elles mises à profit ?

Les normes de qualité et en matière d'éthique devraient éclairer la réflexion subséquente sur la méthodologie, la conception, la mise en œuvre et la gestion du processus d'évaluation.

Pour adapter les critères à une évaluation donnée, il importe également d'analyser en profondeur les liens de causalité et d'évaluer dans quelle mesure l'évaluation parviendra à attribuer des effets à l'intervention évaluée. Ces considérations peuvent aider à gérer les attentes des parties prenantes et à déterminer les critères qui devront être examinés en profondeur. Une telle démarche est particulièrement importante pour les critères d'efficacité et d'impact (voir les deux sections qui suivent), mais est aussi indirectement applicable – et peut servir – à d'autres critères. Par exemple, les bénéfices réels (ou projetés) attribués à l'intervention au titre de l'évaluation de l'efficacité et de l'impact pourraient également servir à l'évaluation de l'efficacité et de la viabilité/durabilité. L'analyse de l'évaluabilité est utile pour s'assurer que la finalité de l'évaluation est parfaitement comprise¹. Le Chapitre 4 contient des explications supplémentaires sur la manière d'interpréter chaque critère.

Encadré 3.2. Définir la finalité de l'évaluation : exemples de l'Ouganda, du Kenya et de la Suède

Évaluation par l'Agence danoise de développement international (Danida) des programmes relatifs à l'eau, à l'assainissement et à l'environnement en Ouganda

Le Département de l'évaluation du ministère danois des Affaires étrangères a commandé une évaluation indépendante des initiatives danoises menées sur la période 1990-2017 en vue de développer les secteurs de l'eau, de l'assainissement et de l'environnement en Ouganda.

Les objectifs de l'évaluation étaient les suivants :

- recenser les résultats et les réalisations dans les sous-secteurs ;
- analyser la « valeur ajoutée » du soutien danois à ces sous-secteurs ;
- dégager des enseignements.

À ces fins, l'évaluation a mis l'accent sur l'évaluation de l'efficacité à deux niveaux : 1) sectoriel et transversal et 2) sous-sectoriel. Elle a également apprécié l'intérêt des initiatives et examiné dans quelle mesure les résultats obtenus étaient durables.

Évaluation conjointe du Haut-commissariat des Nations Unies pour les réfugiés (HCR) et du Danemark

Le but de l'évaluation menée conjointement par le HCR et le Danemark concernant l'action en faveur des réfugiés dans le camp de Kalobeyei, au Kenya, et alentour est défini comme suit :

L'évaluation a pour principal objectif de contribuer à l'apprentissage concernant le modèle de camp intégré mis en œuvre à Kalobeyei et alentour. En recensant les enseignements tirés de la mise en œuvre d'une initiative concrète visant à relier aide humanitaire et aide au développement à long terme, l'intention est d'apporter des informations sur les possibilités et les défis liés à la conception et à la mise en œuvre d'un modèle de solution intégrée.

Évaluation du soutien apporté par l'Agence suédoise de coopération pour le développement international (Asdi) en faveur de la consolidation de la paix dans les situations de conflit et d'après-conflit

Le rapport d'évaluation donne un exemple utile de la manière dont l'évaluation reflète le but visé. Ainsi qu'il l'explique :

L'évaluation a pour but ou utilité envisagée de systématiser les enseignements tirés de la pratique de la consolidation de la paix. Elle servira d'intrant dans le processus de conceptualisation et de formulation de l'approche que l'Asdi appliquera dans ce domaine. Cette approche influera à son tour sur la planification stratégique et la conception des activités futures de soutien de l'Asdi dans les contextes touchés par un conflit. En outre, l'évaluation est censée contribuer à une meilleure compréhension de la consolidation de la paix en tant que concept et pratique.

L'évaluation visait spécifiquement à évaluer comment l'Asdi aborde la problématique de la consolidation de la paix au niveau stratégique dans des contextes différents. À cette fin, elle a accordé une attention particulière à quatre critères :

- la pertinence des activités de consolidation de la paix menées par l'Asdi par rapport aux besoins du contexte et des bénéficiaires, et ses priorités stratégiques ;
- l'efficacité de ses activités de consolidation de la paix en termes de contribution aux objectifs globaux dans ce domaine et son aptitude à offrir un cadre propice aux activités de consolidation de la paix menées par ses partenaires ;
- l'impact de ses activités de consolidation de la paix ;
- la pérennité de ses activités de consolidation de la paix.

Pour l'ensemble de ces quatre critères, un accent particulier a été mis sur les résultats dans les domaines de l'égalité entre les sexes, de l'autonomisation et des droits des femmes, et de l'inclusion des groupes marginalisés et des minorités ethniques. Les évaluations relatives aux quatre études de cas ont porté principalement sur des dimensions de la marginalisation liées à la consolidation de la paix et abordées dans la littérature ainsi qu'au cours des entretiens. La marginalisation des groupes a été principalement considérée sous l'angle de l'appartenance ethnique ou comme une conséquence de l'appartenance à une minorité, les facteurs ethniques constituant la principale cause de conflit dans les quatre cas étudiés.

Source : Danida (2019^[3]), *Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017)*, <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf> ;

ADE (2019^[4]), *Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya*, https://um.dk/en/danida-en/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c-974b-c171fbc2452d ;

Bryld (2019^[5]), *Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report*, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665c57c1/eva2019_5_62214en.pdf

Comprendre la dimension temporelle

Les critères peuvent être appliqués à différents moments dans le temps. Chacun d'entre eux peut être utilisé pour évaluer une intervention avant, pendant ou après sa mise en œuvre. De même, ils peuvent être évalués à différents moments du cycle de vie de l'intervention. Toutefois, l'interprétation des critères et des sources de données peut varier selon le moment choisi. Ainsi, avant une intervention, l'efficacité et la viabilité/durabilité sont de simples projections, tandis qu'après l'intervention, des données plus nombreuses sont disponibles, qui permettent de tirer des conclusions plus robustes.

Les critères – et les questions d'évaluation qui leur sont associées – devraient donc tenir compte des deux principaux aspects temporels suivants : 1) le moment, dans le cycle de vie de l'intervention, où s'effectue l'évaluation et 2) le stade où se trouve l'intervention ou le moment dans la chaîne de résultats sur lequel l'évaluation se concentrera.

Le cadre conceptuel de chaque critère reste le même, que l'évaluation soit effectuée avant, pendant ou après une intervention. En revanche, les données et les autres éléments factuels disponibles pour *évaluer* les critères et les méthodes utilisées changent. Les évaluateurs devraient garder ces différences à l'esprit lorsqu'ils examinent les constatations (potentielles) avec les parties prenantes de l'évaluation, car elles peuvent influencer sur la manière dont l'utilité et la crédibilité de l'évaluation sont perçues. Ainsi, une évaluation *ex ante* de la viabilité/durabilité (avant le démarrage d'une intervention) pourra évaluer la probabilité que les bénéficiaires de l'intervention perdurent dans le temps, en examinant la conception de l'intervention et les données disponibles attestant la validité des hypothèses relatives à la pérennité des bénéficiaires attendus. Une fois l'intervention achevée, l'évaluation de la viabilité visera à déterminer si les bénéficiaires ont effectivement été pérennes, cette fois-ci en s'appuyant sur des données et d'autres éléments factuels tirés des bénéficiaires réellement obtenus.

Les évaluations rétrospectives devront prendre en compte le contexte et les données disponibles au moment en question pour opérer des jugements sur la base d'attentes raisonnables quant aux actions qui auraient pu ou dû être menées. Il serait injuste de juger les actions des concepteurs de programmes achevés au regard des informations disponibles au moment de l'évaluation et non de celles alors connues des intéressés. Cela étant, il a été observé dans le cadre de nombreuses évaluations que les informations disponibles n'avaient pas été pleinement utilisées (là où elles auraient pu l'être), ce qui aurait pu accroître la pertinence de l'intervention. Par exemple, les populations locales n'avaient pas été suffisamment consultées ni associées au stade de la conception de l'intervention. Cette lacune, dont on aurait raisonnablement pu s'attendre à ce qu'elle soit évitée, devrait donc être signalée dans une évaluation.

Utiliser les critères pour améliorer les questions d'évaluation

La formulation de questions d'évaluation adéquates constitue une partie essentielle du processus d'évaluation et l'utilisation des six critères interagit avec la prise de décisions sur ces questions, à laquelle elle contribue². Le processus débute par une réflexion sur la finalité de l'évaluation, la manière dont elle sera mise à profit et par qui. Un dialogue efficace avec les parties prenantes, à travers un processus participatif bien conçu, peut aider les évaluateurs et les responsables à comprendre comment ces parties mettront à profit l'évaluation. Ce dialogue devrait être complété par une compréhension profonde de l'intervention et du contexte dans lequel elle s'inscrit, ainsi que de ses objectifs et de la théorie du changement appliquée. Une phase initiale peut servir à étudier ces aspects et à formuler (en petit nombre) des questions clés auxquelles l'évaluation devra répondre.

Après cette phase initiale, les critères d'évaluation permettent d'opérer des vérifications sous plusieurs angles afin de repérer les éventuelles omissions. Il est ainsi possible de développer et d'affiner les questions – un aspect essentiel pour la conception de l'évaluation. Une telle démarche permet de systématiser le processus et garantit l'exhaustivité de l'évaluation. C'est un élément essentiel d'une « meilleure évaluation ». Il est fortement recommandé d'adopter une approche globale claire et cohérente pour cette phase d'élaboration des questions d'évaluation.

Il se peut que l'institution commanditaire ait pris la décision d'évaluer les interventions selon certains critères. Dans ce cas, l'évaluation devrait être menée conformément aux orientations de l'institution et les critères devraient être appliqués de façon réfléchie, à des fins de cohérence.

Le manuel d'évaluation de l'Asdi en est une illustration, qui fournit des exemples de questions types en relation avec chaque critère (Molund et Schill, 2004^[6]). Les orientations et les notes techniques élaborées

à l'intention des responsables d'évaluation et des évaluateurs du Programme alimentaire mondial constituent un autre exemple de la manière dont les institutions interprètent les critères et en appréhendent les composantes (PAM, 2016^[7]).

Liens entre les critères

Les critères constituent de multiples prismes à travers lesquels une intervention et ses résultats peuvent être considérés. Ils sont liés entre eux dans le sens où les notions qui les sous-tendent permettent d'analyser des dimensions complémentaires du processus d'obtention des résultats. Par exemple, l'efficacité et l'impact examinent des niveaux différents de la chaîne de résultats, selon la manière dont les objectifs de l'intervention ont été définis. Il existe donc une corrélation entre ces deux critères tout au long de la chaîne de causalité.

Les critères sont souvent interdépendants. Ainsi, une intervention ne présentant pas de pertinence au regard des priorités des bénéficiaires a peu de chances de produire l'impact attendu (à moins que celui-ci puisse être assuré par d'autres voies inédites). Une intervention mal mise en œuvre (moins efficace) risque également de se traduire par des effets escomptés. À l'inverse, il peut aussi arriver qu'une intervention hautement pertinente soit inefficace, ou qu'une intervention très cohérente soit particulièrement inefficace en raison de coûts de transaction accrus. Les évaluateurs devraient étudier et analyser les liens et les synergies entre les différents critères, en s'intéressant notamment à la question de savoir s'ils sont liés par un lien de causalité et dans ce cas, de quelle manière.

La plupart des évaluations tirent des conclusions à partir des constatations dressées pour chaque critère ainsi qu'une conclusion générale, et notent parfois la performance à l'aide d'une valeur numérique. Pour tirer des conclusions sur l'intervention, et selon le but de l'évaluation, les évaluateurs devraient examiner la situation dans son ensemble et réfléchir à la manière de pondérer comme il se doit chacun des critères appliqués. À cet égard, certaines institutions définissent parfois un critère dominant (éliminatoire). Si les résultats ne sont pas satisfaisants au regard de ce critère, aussi bons les résultats obtenus au niveau des autres critères soient-ils, l'intervention est considérée comme infructueuse (ou, s'il s'agit d'une évaluation *ex ante*, elle n'est pas financée).

On trouvera un examen de liens plus spécifiques entre chaque critère au Chapitre 4.

Sélectionner les critères à utiliser

Comme indiqué précédemment, la finalité, les priorités, la portée et le contexte de l'intervention et de l'évaluation détermineront l'importance relative accordée à chaque critère. Les évaluateurs devraient considérer la valeur ajoutée relative qu'apportera chaque critère, ce qui suppose de répondre à deux questions fondamentales : Ce critère est-il un élément d'appréciation important pour l'évaluation ? Est-il possible de répondre aux questions relatives à ce critère ?

Si les utilisateurs peuvent être tentés d'appliquer automatiquement l'ensemble des six critères, quel que soit le contexte, la meilleure approche (c'est-à-dire, celle qui est cohérente avec l'intention initiale des critères et qui permettra de mener une évaluation de qualité supérieure) consiste à sélectionner délibérément les critères à appliquer et à les utiliser d'une manière adaptée à l'évaluation et aux questions auxquelles cette dernière vise à répondre.

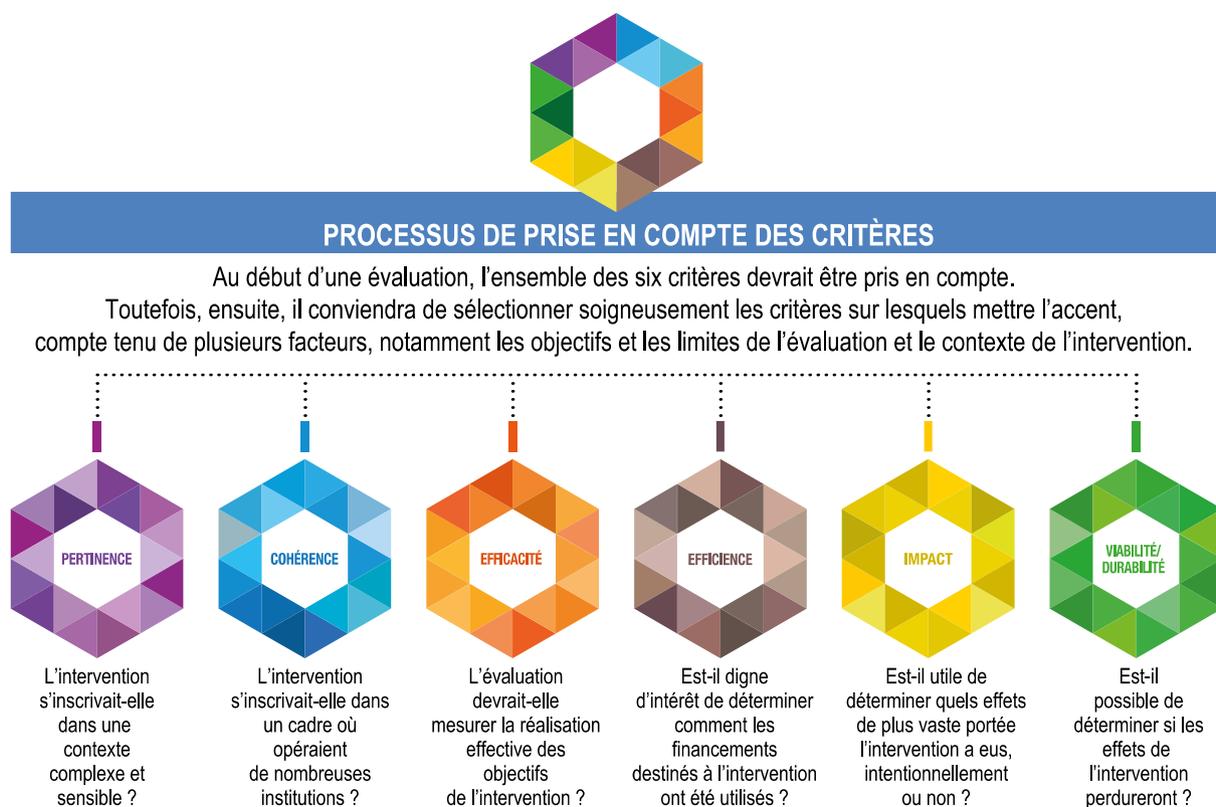
Pour y parvenir, il convient de se poser des questions du type :

- Si une seule question pouvait être posée concernant cette intervention, quelle serait-elle ?

- Quelles questions trouvent leur meilleure réponse en menant une évaluation et lesquelles en recourant à d'autres moyens (par exemple, un projet de recherche, une synthèse de données, un exercice de suivi ou un processus d'apprentissage itératif) ?
- Les données disponibles sont-elles suffisantes pour apporter une réponse satisfaisante à cette question ? À défaut, des données de meilleure qualité ou en plus grand nombre seront-elles disponibles par la suite ?
- Qui a apporté sa contribution pour élaborer la liste de questions ? Est-ce qu'il manque des angles de vue importants ?
- Disposons-nous de suffisamment de temps et de ressources pour prendre en compte de manière adéquate l'ensemble des critères qui présentent un intérêt, ou obtiendrons-nous des informations plus utiles en axant l'analyse sur certains critères seulement ?

Lorsqu'il s'agit d'utiliser les critères, il importe de trouver un équilibre entre flexibilité (éviter toute application mécanique des critères dans leur ensemble) et « picorage » (choisir uniquement les critères les plus faciles ou ceux qui sont les plus susceptibles de déboucher sur des résultats positifs). En particulier, il ne faut pas reculer devant la tâche parfois difficile de répondre à des questions essentielles sur l'impact et la cohérence. Le Graphique 3.1 ci-après présente certains aspects à prendre en compte ; il convient de noter qu'il s'agit d'exemples et non d'une liste exhaustive.

Graphique 3.1. Réfléchir à la manière de prendre en compte les critères pour une évaluation donnée



S'agissant de la conception et de la mise en œuvre de l'évaluation, une bonne connaissance des parties prenantes concernées – par l'intervention mais aussi par l'évaluation – peut permettre de repérer les tensions potentielles entre leurs intérêts et leurs priorités. Ainsi, l'intérêt des bénéficiaires pourra porter

davantage sur l'efficacité (par exemple, le fait de savoir si la participation à un programme de traitement de la malnutrition permet d'améliorer la santé de leurs enfants), tandis que les organismes d'exécution se soucieront plutôt de l'efficacité (dans l'optique de proposer le traitement à un plus grand nombre de familles, par exemple). Dans la plupart des cas, une évaluation ne peut seule répondre à toutes les questions potentiellement intéressantes et des choix doivent être opérés. Pour accroître la probabilité que les questions non couvertes par l'évaluation le soient dans le cadre d'un autre processus, une bonne pratique consiste à consigner les informations portant sur le processus et les résultats des discussions relatives à la hiérarchisation des besoins des différentes parties prenantes et aux décisions sur les questions d'évaluation.

Mettre en œuvre et adapter les critères au niveau institutionnel

Les définitions des critères et les présentes Orientations offrent un socle commun et un ensemble de définitions communément admises, à titre de point de départ. Il est néanmoins indispensable de les adapter au contexte institutionnel. Les évaluateurs et les responsables d'évaluation utilisent souvent les critères pour le compte d'une organisation, d'un ministère ou d'une autre institution en charge du développement, doté(e) de son propre mandat, de ses propres priorités stratégiques, de sa propre politique d'évaluation et de ses propres normes et orientations – autant d'éléments dont décident leurs organes de direction.

Lorsque les évaluateurs et les commanditaires étudient les moyens de mettre en œuvre et d'appliquer les critères, il importe qu'ils tiennent dûment compte, pour prendre leurs décisions, des priorités stratégiques et de la culture de l'organisation ainsi que des possibilités qu'elle offre. Certains termes (par exemple, « impact ») ne sont pas toujours employés de la même manière et il convient de prendre garde au risque de confusion ou de mauvaise interprétation des critères et de leur intention de départ. Une telle démarche aidera les évaluateurs et les commanditaires à appliquer les critères, permettra d'optimiser l'utilisation des constatations de l'évaluation et renforcera la pertinence par rapport aux besoins des utilisateurs cibles. Les responsables devraient encourager les échanges entre évaluateurs, commanditaires et public visé par l'évaluation de façon à examiner la manière dont les critères devraient être appliqués et interprétés. Un tel processus peut permettre de concevoir des évaluations crédibles et opportunes qui répondent mieux aux besoins des utilisateurs.

Les exigences d'une institution en ce qui concerne les méthodologies à suivre peuvent également avoir une incidence sur la manière dont les critères sont appliqués. Les évaluateurs devraient se référer aux exigences et orientations spécifiques de leur organisation/commanditaire. D'autres sources pertinentes telles que les orientations du Groupe des Nations Unies pour l'évaluation (GNUE) ou les normes de bonnes pratiques du Groupe de coopération pour l'évaluation (ECG) et les orientations du Réseau pour l'apprentissage actif sur la responsabilité et la performance au sein de l'action humanitaire (Active Learning Network for Accountability and Performance in Humanitarian Action, ALNAP) sur l'évaluation dans les contextes humanitaires sont également très utiles lorsqu'elles sont applicables.

Répondre aux exigences du Programme de développement durable à l'horizon 2030 et des Objectifs de développement durable

Outre la nécessité d'adapter les critères aux institutions qui les utilisent, la manière dont les critères sont compris et appliqués doit refléter le contexte général, ce qui influe sur leur utilisation par les responsables d'évaluation, les évaluateurs et les parties prenantes. Pour les dix prochaines années, ce sont les Objectifs de développement durable (ODD) et le Programme de développement durable à l'horizon 2030 qui constitueront le cadre d'action global et l'ensemble d'objectifs mondiaux les plus importants, en particulier pour les évaluateurs de la coopération internationale pour le développement.

Les principaux éléments du Programme 2030 sont notamment les suivants :

- l'accès universel aux fruits du développement ;
- l'inclusion, en particulier des personnes qui risquent le plus d'être laissées de côté ;
- les droits humains, l'égalité femmes-hommes et d'autres considérations d'équité ;
- la durabilité environnementale, le changement climatique et la gestion des ressources naturelles ;
- la complexité du contexte et des interventions à l'appui du développement ;
- les synergies entre les acteurs engagés dans le processus de développement.

Ce cadre influence à la fois les interventions et leur évaluation, y compris la manière dont les critères sont interprétés, ainsi que le processus d'évaluation proprement dit (notamment les personnes qui sont chargées d'appliquer les critères et de définir des questions prioritaires). L'Encadré 3.3 donne des orientations sur la façon dont le Programme 2030 peut être utilisé pour définir des programmes d'évaluation nationaux et l'Encadré 3.4 montre comment le système allemand d'évaluation du développement y est parvenu. Des efforts analogues ont été déployés par plusieurs États – dont le Costa Rica, le Nigéria et la Finlande – et par des organisations non gouvernementales (ONG), qui livrent des enseignements utiles pour les évaluateurs et les responsables d'évaluation (D'Errico, Geoghe et Piergallini, 2020^[9]).

Les définitions révisées des critères prennent également en compte ces éléments à plusieurs égards, par exemple en accordant une importance particulière au contexte et aux points de vue et priorités des bénéficiaires pour examiner la pertinence, l'efficacité et l'impact ; en tenant compte de l'équité des résultats pour analyser l'efficacité et l'impact ; et en adoptant une approche intégrée pour évaluer la cohérence. Les principes directeurs tiennent également compte du Programme 2030 en prônant un mode de réflexion intégré.

Encadré 3.3. S'appuyer sur les ODD pour définir des programmes d'évaluation nationaux

La publication intitulée « L'évaluation pour relier les priorités nationales aux ODD : Un guide pour les commanditaires et responsables d'évaluation » expose cinq aspects à prendre en compte pour aider les pays à concevoir des programmes d'évaluation et à renforcer l'utilité des évaluations. Les évaluateurs sont encouragés à appliquer la « théorie des systèmes complexes » et à :

- penser au-delà des politiques, programmes et projets individuels ;
- examiner les forces au niveau macro qui influent sur les chances de réussite ou d'échec ;
- avoir une compréhension nuancée de la notion de « réussite » ;
- reconnaître l'importance de la culture ;
- faire leur l'approche de réflexion évaluative et la gestion adaptative.

Source : Ofir et al. (2016^[9]), *Briefing: Five considerations for national evaluation agendas informed by the SDGs*, <https://pubs.iied.org/sites/default/files/pdfs/migrate/17374IIED.pdf>

Encadré 3.4. Relier les critères au Programme 2030 et aux ODD

Les questions qui suivent ont été élaborées par le ministère fédéral allemand de la Coopération économique et du développement (BMZ), avec le concours de l'Institut allemand pour l'évaluation du développement (DEval), afin d'aider les évaluateurs à déterminer quelle contribution globale a été apportée à la réalisation du Programme 2030 et des ODD. Chaque question correspond à des principes consacrés par les ODD et se rapporte également aux critères.

Universalité, responsabilité commune et redevabilité

- Dans quelle mesure l'intervention contribue-t-elle à la réalisation des ODD ? (voir critère d'impact)
- Dans quelle mesure l'intervention est-elle conçue de manière à utiliser les structures et les systèmes existants (des partenaires/d'autres donateurs/d'organisations internationales) pour mettre en œuvre les activités et dans quelle mesure ces structures et systèmes sont-ils utilisés ? (voir critère de cohérence)
- Les tâches sont-elles partagées avec d'autres donateurs et partenaires au développement pour mettre en œuvre l'intervention ? Si oui, dans quelle mesure ? (voir critère de cohérence)
- Dans quelle mesure les systèmes communs sont-ils utilisés pour le suivi, l'apprentissage et l'exercice de la redevabilité ? (voir critère de cohérence)

Interaction entre les dimensions économique, environnementale et sociale du développement

- Dans quelle mesure l'intervention adopte-t-elle une approche globale du développement durable (dimensions sociale, environnementale et économique) ? (voir critère de pertinence)
- Dans quelle mesure existait-il des interactions positives ou négatives, intentionnelles ou non, entre les réalisations sociales, économiques et environnementales, et quel a été l'impact global de l'intervention ? (voir critères d'efficacité et d'impact)
- Comment l'intervention a-t-elle contribué à favoriser les interactions positives ou négatives, intentionnelles ou non, entre les réalisations sociales, économiques et environnementales, et quel a été son impact global ? (voir critères d'efficacité et d'impact)

Inclusivité

- Dans quelle mesure l'intervention est-elle cohérente avec les normes et les critères internationaux relatifs à la participation et à la promotion des groupes particulièrement défavorisés et vulnérables ? (voir critère de cohérence)
- Dans quelle mesure y a-t-il eu des évolutions générales sur le plan du développement, positives ou négatives, intentionnelles ou non, au niveau des groupes particulièrement défavorisés et vulnérables (différenciation possible selon l'âge, le revenu, le sexe, l'origine ethnique, etc.) ? (voir critère d'impact)
- Comment l'intervention a-t-elle contribué à favoriser les évolutions générales sur le plan du développement, positives ou négatives, intentionnelles ou non, au niveau des groupes particulièrement défavorisés et vulnérables (différenciation possible selon l'âge, le revenu, le sexe, l'origine ethnique, etc.) ? (voir critère d'impact)
- Dans quelle mesure l'intervention a-t-elle contribué au renforcement de la résilience des groupes particulièrement défavorisés ou vulnérables (différenciation possible selon l'âge, le revenu, le sexe, l'origine ethnique, etc.) ? (voir critère de viabilité/durabilité)

Source : BMZ et DEval (2020^[10]), *Evaluation Criteria for German Bilateral Development Co-operation*

Appliquer un prisme sexospécifique aux critères

Les évaluateurs devraient adopter des méthodes de travail qui prennent en compte de manière réfléchie les différences d'expérience et d'impact selon le sexe, ainsi que leur interaction avec d'autres formes de discrimination dans un contexte donné (fondées, par exemple, sur l'âge, la « race » et l'origine ethnique, le statut social). Quelle que soit l'intervention, les évaluateurs devraient s'intéresser à la manière dont les dynamiques de pouvoir fondées sur le sexe se croisent et interagissent avec d'autres formes de discrimination d'une manière qui influe sur la mise en œuvre et les résultats de l'intervention. Il peut s'agir d'étudier la façon dont l'économie politique et le contexte socioculturel dans lequel s'inscrivent les interventions influencent la poursuite et la réalisation des objectifs.

Le fait d'appliquer un prisme sexospécifique peut livrer des données factuelles à l'appui de l'apprentissage et de l'exercice de la redevabilité, tout en favorisant la réalisation des objectifs liés à l'égalité des sexes. Pour ce faire, il convient notamment de prendre les mesures concrètes suivantes :

- adopter, pour les évaluateurs, les responsables et les commanditaires, des méthodes de travail inclusives qui favorisent une participation appropriée à la prise de décisions, à la collecte de données, à l'analyse et à la diffusion des constatations ;
- examiner la mesure dans laquelle, dans le cadre de l'intervention, le genre interagit avec d'autres barrières sociales pour compromettre l'égalité des chances ;
- examiner la manière dont une intervention interagit avec le contexte législatif, économique, politique, religieux et socioculturel pour mieux interpréter les différences en termes d'expérience des parties prenantes et en termes d'impact ;
- prendre en compte les constructions sociales de la masculinité et de la féminité ainsi que toute évolution de la dynamique et des rôles entre les sexes ;
- lors de la sélection des évaluateurs, analyser leurs compétences en matière d'approches d'évaluation tenant compte des sexospécificités ainsi que leur expérience acquise dans des contextes de travail diversifiés.

Le tableau ci-après vise à aider les évaluateurs à réfléchir sur la manière dont ils peuvent appliquer un prisme sexospécifique aux critères :

Tableau 3.1. Appliquer un prisme sexospécifique aux critères

Critères	Questions pour guider les évaluateurs
Pertinence	<p>L'intervention a-t-elle été conçue de manière à répondre aux besoins et aux priorités des différents sexes ? Si oui, comment ?</p> <p>Dans quelle mesure la conception de l'intervention tient-elle compte des droits des personnes de tout sexe et prévoit un retour d'information de diverses parties prenantes locales, y compris les groupes marginalisés ?</p> <p>L'intervention répond-elle aux besoins concrets et stratégiques des différents sexes ?</p>
Cohérence	<p>Dans quelle mesure la conception, la mise en œuvre et les résultats de l'intervention sont-ils cohérents avec la législation et les engagements internationaux en faveur de l'égalité et des droits femmes-hommes, notamment la Convention des Nations Unies sur l'élimination de toutes les formes de discrimination à l'égard des femmes (CEDAW), la Déclaration et le Programme d'action de Pékin, le Programme d'action de la Conférence internationale sur la population et le développement et le Programme de développement durable à l'horizon 2030 ?</p> <p>Dans quelle mesure l'intervention soutient-elle la législation et les initiatives nationales visant à améliorer l'égalité des sexes et les droits de la personne ? Quels enseignements peuvent être tirés ?</p>
Efficacité	<p>L'intervention a-t-elle atteint ses objectifs et résultats attendus d'une manière qui contribue à l'égalité des sexes ? Si oui, comment ?</p> <p>Les résultats ont-ils été différents selon le groupe de personnes ? Si oui, comment et pourquoi ? Des approches différentes ont-elles été nécessaires pour atteindre les personnes de différents sexes ? Le suivi et l'analyse des effets différenciés ont-ils été suffisants ? L'intervention a-t-elle été adaptée pour répondre aux éventuelles préoccupations et maximiser son efficacité ?</p> <p>La théorie du changement et le cadre de résultats reposaient-ils sur une analyse de la situation en matière d'égalité entre les sexes, sur une analyse de l'économie politique et sur les droits humains ? Si oui, dans quelle mesure ?</p> <p>Dans quelle mesure l'efficacité varie-t-elle selon le sexe des personnes et pourquoi ?</p>
Efficience	<p>Des ressources différentes ont-elles été allouées d'une manière tenant compte de l'égalité entre les sexes ? Si oui, de quelle manière ont-elles été allouées ? L'allocation différenciée des ressources était-elle appropriée ?</p> <p>Les coûts d'investissement par personne ciblée sont-ils en adéquation avec les besoins différenciés des personnes de sexes différents ?</p>
Impact	<p>Les impacts ont-ils été les mêmes pour les différents sexes ou y avait-il des différences en termes d'engagement, d'expérience et d'impact ? Si oui, comment s'expliquent ces différences d'impact ?</p> <p>Dans quelle mesure les impacts liés au sexe se croisaient-ils avec d'autres barrières sociales, notamment la « race »/l'origine ethnique, le handicap, l'âge et l'orientation sexuelle, de manière à contribuer à des différences en termes d'expérience et de réalisation ?</p> <p>Comment les normes et les barrières sexuées en vigueur dans le contexte politique, économique, religieux, législatif et socioculturel plus vaste ont-elles influé sur les réalisations ?</p> <p>Dans quelle mesure l'impact a-t-il contribué à des rapports de force égaux entre les personnes de sexe différent ainsi qu'à l'évolution des normes et des systèmes sociaux ?</p>
Viabilité/durabilité	<p>L'intervention a-t-elle contribué à l'égalité entre les sexes au sein des systèmes juridique, politique, économique et social plus vastes ? Si oui, de quelle manière et dans quelle mesure ? Cela s'est-il traduit par des changements durables au niveau de normes sociales préjudiciables pour les personnes de tous sexes ou d'un sexe donné ? Si non, pourquoi pas ?</p> <p>Les réalisations en matière d'égalité entre les sexes perdureront-elles une fois l'intervention achevée ? Les processus ont-ils contribué à pérenniser ces acquis ? Des mécanismes ont-ils été mis en place pour promouvoir la réalisation de l'égalité entre les sexes sur le long terme ?</p>

Utilisation d'autres critères

Les six critères ont vocation à former un ensemble complet reflétant pleinement toutes les notions importantes qui devraient être couvertes par les évaluations. S'ils sont appliqués de manière réfléchie et

en tenant compte du contexte, ils conviendront pour des évaluations dans le domaine du développement durable comme de l'action humanitaire.

Pour autant, dans certains contextes, d'autres critères sont utilisés. Ainsi, dans leurs politiques d'évaluation, de nombreuses institutions commandent des analyses de domaines d'intérêt particuliers. En 2020, une évaluation des programmes menés par l'Italie en Bolivie dans le secteur de la santé a recouru à neuf critères, à savoir : la pertinence, l'efficacité, l'efficience, l'impact, la viabilité/durabilité, la cohérence, la valeur ajoutée de la coopération italienne, la visibilité de la coopération italienne et l'appropriation (Eurecna Spa, 2020^[11]). Dans un autre exemple, il s'agissait d'appliquer les critères dans des situations humanitaires, où des critères comme l'adéquation, la portée et l'existence de liens sont particulièrement pertinents³. L'additionnalité est un autre critère qui est parfois appliqué, souvent en relation avec le financement mixte, les financements non souverains et le financement climatique. Différentes définitions de ce terme, y compris différents types d'additionnalité financière et non financière, sont utilisés⁴. Selon la définition appliquée, l'additionnalité peut être évaluée au regard du critère de pertinence, d'efficacité ou d'impact. Elle est parfois elle-même considérée comme un critère transversal.

Les utilisateurs devraient faire preuve de prudence lorsqu'ils envisagent d'ajouter des critères, car une telle démarche peut créer la confusion et conduire à une évaluation trop générale (et donc à une analyse moins utile). S'appuyer sur un nombre limité de critères est utile pour permettre une analyse suffisamment approfondie et garantir la clarté conceptuelle – un point souligné à de nombreuses reprises au cours des consultations menées en vue de l'actualisation des définitions, en 2017-19.

Si l'on décide de recourir à d'autres critères, il importe alors de les définir. Justifier leur ajout peut permettre de comprendre comment ils se situent par rapport aux six critères décrits dans le présent document. Pour favoriser l'apprentissage d'une évaluation à l'autre, il est impératif d'évaluer les mêmes notions ou éléments au regard des mêmes critères.

Quels que soient les critères utilisés, les orientations et les principes fondamentaux décrits dans ce document devraient être appliqués.

Références

- ADE (2019), *Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya*, HCR et DANIDA, https://um.dk/en/danida-en/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c-974b-c171fbc2452d (consulté le 11 janvier 2021). [4]
- Bamberger, M., J. Vaessen et E. Raimondo (2015), *Dealing With Complexity in Development Evaluation - A Practical Approach*, https://www.betterevaluation.org/en/resources/dealing_with_complexity_in_development_evaluation (consulté le 11 janvier 2021). [13]
- BMZ et DEval (2020), *Evaluation Criteria for German Bilateral Development Co-operation*. [10]
- Bryld, E. (2019), *Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report*, Asdi, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf (consulté le 11 janvier 2021). [5]
- D'Errico, S., T. Geoghe et I. Piergallini (2020), *Evaluation to connect national priorities with the SDGs*, IIED, <https://pubs.iied.org/17739IIED> (consulté le 22 février 2021). [8]

- Danida (2019), *Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017)*, Evaluation Department, Ministère des Affaires étrangères du Danemark, <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf> (consulté le 11 janvier 2021). [3]
- Davis, R. (2013), « Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations », n° 40, DIFD, <https://www.gov.uk/government/publications/planning-evaluability-assessments> (consulté le 12 janvier 2021). [12]
- Eurecna Spa (2020), *Bolivia - Evaluation of Health Initiatives (2009-2020)*, Ministère des Affaires étrangères et de la Coopération internationale de l'Italie, http://www.oecd.org/derec/italy/evaluation-report-of-health-initiatives-in-Bolivia-2009_2020.pdf (consulté le 11 janvier 2021). [11]
- Helle, E. et al. (2018), *From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations*, Agence norvégienne de coopération pour le développement (NORAD), https://www.norad.no/globalassets/filer-2017/evaluating/1.18-from-donor-to-partners/1.18-from-donors-to-partners_main-report.pdf (consulté le 11 janvier 2021). [2]
- Molund, S. et G. Schill (2004), *Looking Back, Moving Forward Sida Evaluation Manual*, Asdi, <https://www.oecd.org/derec/sweden/35141712.pdf> (consulté le 11 janvier 2021). [6]
- OCDE (2019), *Des meilleurs critères pour des meilleurs évaluations : Définitions adaptées et principes d'utilisation*, Réseau du CAD sur l'évaluation du développement, Éditions OCDE, Paris, <https://www.oecd.org/fr/cad/evaluation/criteres-adaptees-evaluation-dec-2019.pdf> (consulté le 11 janvier 2021). [1]
- OCDE (2002), *L'évaluation et l'efficacité de l'aide n° 6 - Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats (en anglais, français et espagnol)*, Éditions OCDE, Paris, <https://doi.org/10.1787/9789264034921-en-fr>. [14]
- Ofir, Z. et al. (2016), *Briefing: Five considerations for national evaluation agendas informed by the SDGs*, IIED, Londres, <https://doi.org/10.3138/cjpe.30.3.02.11>. [9]
- PAM (2016), *Technical Note: Evaluation Methodology*, DEQAS, Programme alimentaire mondial, <https://docs.wfp.org/api/documents/704ec01f137d43378a445c7e52dcf324/download/> (consulté le 11 janvier 2021). [7]

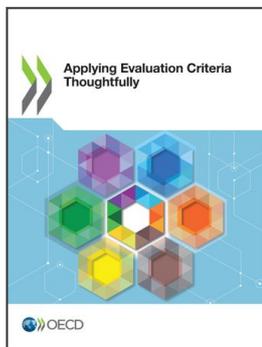
Notes

¹ Par « évaluabilité » est entendu la mesure dans laquelle une activité ou un programme est susceptible d'être évalué(e) d'une manière fiable et crédible. L'analyse de l'évaluabilité nécessite d'examiner à un stade précoce une activité proposée afin de déterminer si ses objectifs sont correctement définis et ses résultats vérifiables (OCDE, 2002^[14]). Voir également Davis (2013^[12]).

² Le processus d'élaboration des questions d'évaluation et de dialogue avec les parties prenantes, en tenant compte également de la complexité de l'intervention, est examiné plus avant, par exemple, dans Bamberger, Vaessen et Raimondo (2015^[13]).

³ Le réseau ALNAP a entrepris d'actualiser ses orientations de 2006 sur l'utilisation des critères dans les contextes humanitaires, en complément de son guide complet sur les évaluations relatives à l'action humanitaire (*Evaluation of Humanitarian Action Guide*).

⁴ Le groupe de travail d'EvalNet sur l'évaluation du financement mixte a contribué aux recherches sur les définitions de l'additionnalité et de notions connexes. Ses conclusions ont été publiées en 2021.



Extrait de :
Applying Evaluation Criteria Thoughtfully

Accéder à cette publication :
<https://doi.org/10.1787/543e84ed-en>

Merci de citer ce chapitre comme suit :

OCDE (2023), « Utiliser les critères d'évaluation dans la pratique », dans *Applying Evaluation Criteria Thoughtfully*, Éditions OCDE, Paris.

DOI: <https://doi.org/10.1787/aff282cc-fr>

Ce document, ainsi que les données et cartes 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. Des extraits de publications sont susceptibles de faire l'objet d'avertissements supplémentaires, qui sont inclus dans la version complète de la publication, disponible sous le lien fourni à cet effet.

L'utilisation de ce contenu, qu'il soit numérique ou imprimé, est régie par les conditions d'utilisation suivantes :
<http://www.oecd.org/fr/conditionsdutilisation>.