

INTRODUCTION À L'ÉVALUATION D'IMPACT

De Patricia J. Rogers
RMIT University (Australie) et BetterEvaluation



Ceci est la première d'une série de quatre notes sur l'évaluation d'impact, élaborées par InterAction avec l'appui financier de la Fondation Rockefeller. Les autres notes de cette série sont : *Faire le lien entre le suivi-évaluation et l'évaluation d'impact* ; *Introduction aux méthodes mixtes dans l'évaluation d'impact* et *Utilisation des résultats d'évaluations d'impact*. On peut trouver la série complète sur les site Internet d'InterAction : <http://www.interaction.org/impact-evaluation-notes>.

EMERCIEMENTS

Cette note d'orientation a bénéficié du feedback du Groupe de travail d'InterAction sur l'évaluation et l'efficacité des programmes. Les informations sur les méthodes d'évaluation dans la Note d'orientation 1 ont été tirées de BetterEvaluation, une collaboration internationale pour le partage d'informations visant à améliorer l'évaluation. On trouvera plus d'informations sur chacune des méthodes sur le site internet : www.betterevaluation.org.

Photo: David Darg

Table des matières

1. Qu'entend-on par « évaluation d'impact » ?	2
2. Pourquoi faudrait-il faire une évaluation d'impact ?	3
3. À quelles questions l'évaluation d'impact veut-elle répondre ?	3
4. Qui devrait mener une évaluation d'impact ?	3
5. Comment choisir les méthodes d'évaluation d'impact ?	5
6. Déterminer les valeurs pour une évaluation d'impact	6
7. Élaborer une théorie ou un modèle de la manière dont l'intervention est censée fonctionner	7
8. Mesurer ou décrire les impacts (et autres variables importantes)	8
9. Expliquer dans quelle mesure les résultats observés ont été produits par l'intervention	10
10. Synthétiser les données probantes	14
11. Rendre compte des constatations et en encourager l'utilisation	14
12. Quand faudrait-il faire une évaluation d'impact ?	15
13. Qu'est-ce qui est nécessaire pour une évaluation d'impact de qualité ?	16
14. Difficultés courantes dans l'évaluation d'impact dans le développement	18
Résumé	19
Références et autres ressources utiles	19

INTRODUCTION

Une évaluation d'impact crédible et appropriée peut rendre le développement plus efficace. L'importance accordée à l'évaluation d'impact dans le développement a soulevé de nombreuses questions. Qu'est-ce qui caractérise une évaluation d'impact crédible et appropriée ? Comment les évaluations d'impact devraient-elles être gérées ? Quelles mesures et sources de données sont adaptées ? Que faudrait-il faire pour promouvoir une bonne utilisation des évaluations d'impact ? Quelles sont les implications de l'intérêt croissant porté à l'évaluation d'impact pour d'autres types de suivi-évaluation (S&E) ?

InterAction a produit une série de notes d'orientation dédiées à ces questions pour aider le personnel de la gestion, des programmes et du S&E dans les ONG internationales à planifier, concevoir, gérer, mener et utiliser des évaluations d'impact. Ces notes peuvent aussi leur fournir des informations qui leur serviront dans des discussions avec des évaluateurs externes, des partenaires et de financeurs.

Cette première note d'orientation, *Introduction à l'évaluation d'impact*, est une présentation générale de l'évaluation d'impact. Elle explique en quoi l'évaluation d'impact est différente d'autres types d'évaluation, et les complète, pourquoi une évaluation d'impact devrait être faite, quand et par qui. Elle décrit différentes méthodes,

approches et conceptions pouvant être utilisées pour les différents aspects de l'évaluation d'impact : préciser quelles sont les valeurs pour l'évaluation, élaborer une théorie de la manière dont l'intervention est censée fonctionner, mesurer ou décrire les impacts et autres variables importantes, expliquer pourquoi des impacts se sont produits, synthétiser les résultats, rendre compte et appuyer l'utilisation. Cette note examine ce qui est considéré comme une bonne évaluation d'impact, une évaluation qui maintient un équilibre entre des impératifs concurrents : être utile, rigoureux, éthique et pratique, et décrit la manière d'y arriver. Tout au long du document, des notes de bas de page donnent les références de lectures complémentaires dans des domaines particuliers.

1. Qu'entend-on par « évaluation d'impact » ?

L'évaluation d'impact étudie les changements causés par une intervention. On peut faire une évaluation d'impact pour des interventions de toutes tailles : un petit projet local de VIH/Sida ; le programme d'une ONG pour le renforcement de la société civile dans son ensemble ; une série de projets de gestion des ressources naturelles mis en œuvre dans une région géographique ; ou un ensemble d'activités mises en œuvre simultanément par différentes organisations pour améliorer les capacités d'une communauté.

Les résultats prévus d'une intervention représentent une grande partie de l'évaluation d'impact, mais il est aussi important d'étudier les résultats imprévus. Dans cette note d'orientation, les impacts sont définis de la manière suivante :

les effets positifs et négatifs, prévus et imprévus, directs et indirects, primaires et secondaires, produits par une intervention (Définition du Comité d'assistance au développement de l'OCDE)¹

Les impacts sont généralement considérés comme se produisant plus tard que les effets immédiats dont ils résultent. Par exemple les effets intermédiaires de l'amélioration d'accès à la terre et de l'augmentation de la participation aux décisions communautaires seront atteints avant l'impact final d'une amélioration de la santé et du bien-être

¹ Les impacts sont parfois définis de manière tout à fait différente. Par exemple, le *Logic Model Development Guide* de la Fondation W.K.Kellogg (www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx) fait la distinction entre les impacts selon qu'ils s'étendent à des personnes autres que celles qui sont directement impliquées dans le programme. Les changements particuliers dans le comportement, les connaissances, les compétences, l'état et le niveau de fonctionnement des participants sont appelés « effets » et seuls les changements aux organisations, communautés ou systèmes résultant des activités du programme dans les sept à dix ans sont décrits comme des « impacts ».

des femmes auquel ils contribueront. La distinction entre les effets et les impacts peut être relative et dépend des objectifs fixés pour une intervention.

Dans la pratique, il est souvent utile d'inclure aussi bien les effets que les impacts dans une évaluation. Cela indique plus rapidement si une intervention fonctionne ou pas – et si elle ne fonctionne pas, cela permet de déterminer où et peut-être pourquoi.

Dans cette note d'orientation, une évaluation d'impact peut être n'importe quelle évaluation qui étudie de manière systématique et empirique les impacts produits par une intervention. Certaines personnes et certaines organisations utilisent une définition plus restreinte de l'évaluation d'impact et ne considèrent que les évaluations contenant un scénario contrefactuel (une estimation de ce qui se serait passé si l'intervention n'avait pas eu lieu), ou un élément contrefactuel d'un type particulier (par exemple une comparaison avec un groupe qui n'a pas bénéficié de l'intervention). L'USAID, par exemple, utilise la définition suivante : « Les évaluations d'impact mesurent le changement dans un effet de développement qui est attribuable à une intervention définie ; les évaluations d'impact sont basées sur des modèles de cause et d'effet et demandent un scénario contrefactuel crédible et rigoureusement défini pour contrôler les facteurs autres que l'intervention qui pourraient expliquer les changements observés. » Ces différentes définitions sont importantes quand on décide quelles méthodes ou types de recherche seront considérés comme crédibles par les utilisateurs prévus de l'évaluation ou par les partenaires ou les financeurs.

L'évaluation d'impact n'est évidemment pas le seul type d'évaluation qui permette un développement effectif. Il est important de s'assurer que l'on n'investit pas dans l'évaluation d'impact (en

termes de temps et d'argent) au détriment du suivi ou d'autres types d'évaluation, comme l'évaluation des besoins, l'évaluation du processus et l'évaluation coûts-bénéfice, qui sont aussi nécessaires pour fournir des informations permettant de prendre des décisions pratiques et politiques. La Note d'orientation 2 examine la manière dont l'évaluation d'impact et ces autres types de suivi-évaluation peuvent être faits de manière à s'appuyer les uns les autres. Par exemple, les données du suivi peuvent être une bonne base pour une évaluation d'impact et une évaluation d'impact peut guider l'élaboration de systèmes de suivi. L'évaluation d'impact fournit les informations nécessaires pour les évaluations de coût-bénéfice et de coût-efficacité.

2. Pourquoi faudrait-il faire une évaluation d'impact ?

La meilleure manière de faire une évaluation d'impact particulière dépend en partie de son objectif et de qui sont ses principaux utilisateurs prévus. Voici quelques raisons courantes de faire une évaluation d'impact :

- **Pour décider s'il faut financer une intervention** – une « évaluation à priori » est faite *avant la mise en œuvre d'une intervention*, pour estimer ses impacts probables et orienter les décisions de financement.
- **Pour décider s'il faut ou non continuer ou développer une intervention.**
- **Pour apprendre comment répliquer ou augmenter une intervention pilote.**
- **Pour apprendre comment réussir à adapter une intervention réussie à une autre situation.**
- **Pour réassurer les financeurs, y compris les bailleurs de fonds et les contribuables (rendre des comptes en amont) que l'argent**

est investi sagement – et que l'organisation apprend ce qui fonctionne ou pas et utilise ces informations pour améliorer les décisions de mise en œuvre et d'investissement à l'avenir.

- **Pour informer les bénéficiaires prévus et les communautés (rendre des comptes en aval) sur la manière dont le programme bénéficie ou non à la communauté et des quelles manières.**

La Note d'orientation 4 montre plus en détail comment appuyer ces différentes manières d'utiliser l'évaluation d'impact.

3. À quelles questions l'évaluation d'impact veut-elle répondre ?

Une évaluation d'impact devrait essayer de répondre seulement à un petit nombre (cinq à sept) de questions clés. L'évaluation s'intéresse à des questions de haut niveau, pas à des questions particulières qui pourraient être posées dans un entretien ou un questionnaire. Il vaut mieux se limiter à un petit nombre de questions en lien direct avec le but plutôt que d'émettre les ressources de l'évaluation, et l'intérêt des utilisateurs, sur un grand nombre de questions.

4. Qui devrait mener une évaluation d'impact ?

Une évaluation d'impact peut être menée par : un évaluateur externe ou une équipe d'évaluation externe ; une unité interne mais séparée de l'organisation de mise en œuvre ; les personnes impliquées dans l'intervention (y compris les membres de la communauté) ; ou une équipe combinée comprenant des évaluateurs internes ou externes.

Un évaluateur externe peut offrir une expertise et une expérience qui n'existent peut-être pas à

Exemples de questions clés pour l'évaluation d'impact

Impact général

- Est-ce que cela a fonctionné ? Est-ce que [l'intervention] a produit [les impacts prévus] à court, moyen et long terme ?
- Pour qui, de quelles manières et dans quelles circonstances [l'intervention] a-t-elle fonctionné ?
- Quels impacts imprévus (positifs et négatifs) [l'intervention] a-t-elle produit ?

Nature et distribution des impacts

- Les impacts ont-ils des chances d'être durables ?
- Ces impacts ont-ils atteint tous les bénéficiaires prévus ?

Influence d'autres facteurs sur les impacts

- Comment [l'intervention] s'est-elle associée avec d'autres interventions, programmes ou services pour parvenir aux effets ?

- Qu'est-ce qui a aidé [l'intervention] à atteindre ces impacts ou qu'est-ce qui l'en a empêchée ?

Comment cela fonctionne

- Comment [l'intervention] a-t-elle contribué aux [impacts prévus] ?
- Quelles étaient les caractéristiques particulières de [l'intervention] qui ont fait une différence ?
- Quelles variations y a-t-il eu dans la mise en œuvre ?
- Quelle a été la qualité de la mise en œuvre sur les différents sites ?
- Dans quelle mesure les différences d'impact peuvent-elles être expliquées par les variations dans la mise en œuvre ?

Correspondance entre les impacts prévus et les besoins

- Dans quelle mesure les impacts correspondent-ils aux besoins des bénéficiaires prévus ?

l'intérieur de l'organisation. Il peut aussi être plus indépendant et crédible qu'un évaluateur interne. Par exemple, la politique d'évaluation de l'USAID exige que la plupart des évaluations soient faites par un évaluateur externe.

Cependant, les évaluateurs externes ne semblent pas toujours impartiaux à certaines parties prenantes, car leur collecte et leur interprétation des données peuvent être affectées par leur manque de familiarité avec la situation. Dans certains cas, l'implication de parties prenantes du programme et/ou de membres de la communauté peut ajouter rigueur et crédibilité en permettant un meilleur accès aux données (surtout aux informateurs clés) et une interprétation plus appropriée des données.

En particulier, les pratiques suivantes peuvent souvent donner les meilleures évaluations : établir une équipe d'évaluateurs avec des perspectives externes et internes ; garantir la transparence pour les données qui sont utilisées dans l'évaluation et la manière dont elles sont utilisées ; la triangulation, en utilisant plusieurs sources de données probantes (qui se complètent) et plusieurs perspectives pour l'analyse et l'interprétation. Il est particulièrement utile d'avoir dans l'équipe des experts locaux en évaluation qui connaissent la situation, l'histoire et des interventions comparables d'autres agences.

Une évaluation peut être gérée par un groupe interne (peut-être un comité de pilotage interne, informé par un groupe consultatif ayant des

membres externes) ou par un groupe combiné. Généralement, les approches participatives pour la gestion des évaluations impliquent le personnel du programme, des membres de la communauté et des partenaires de développement. Ils ne participent pas seulement à la collecte des données mais aussi à la négociation de l'objectif de l'évaluation d'impact, à la préparation des principales questions d'évaluation, à l'élaboration d'une évaluation qui répondra à ces questions et au suivi des résultats.²

5. Comment choisir les méthodes d'évaluation d'impact ?

Il y a eu énormément de débats sur l'évaluation dans le développement et plus généralement sur les meilleures méthodes pour l'évaluation d'impact. Ces discussions reflètent des visions différentes de ce qui constitue une donnée probante crédible, rigoureuse et utile et de qui devrait mener et contrôler les évaluations.

Certaines organisations et certains évaluateurs considèrent qu'il faudrait utiliser autant que possible des méthodes ou des conceptions de recherche particulières, par exemple des essais aléatoires contrôlés ou des méthodes participatives. D'autres considèrent qu'il faut utiliser des méthodes appropriées à la situation. C'est-à-dire choisir des méthodes qui correspondent à l'objectif de l'évaluation, au type de questions

posées, aux ressources disponibles et à la nature de l'intervention, en particulier si elle est standardisée ou adaptative, et si les interventions fonctionnent plus ou moins de la même façon partout et pour tous ou si le contexte les affecte particulièrement.

Quand on choisit des méthodes, il est important de traiter chacun des six aspects d'une évaluation d'impact :

- **Préciser quelles sont les valeurs pour l'évaluation** : qu'est-ce qui sera, ou ne sera pas, considéré comme souhaitable pour ce qui est des processus, des impacts et de la distribution des bénéfices ?
- **Élaborer et/ou tester une théorie** de la manière dont l'intervention est censée fonctionner : on parle parfois de théories du changement, de modèles logique ou de théories de programme.
- **Mesurer ou décrire** les impacts et autres variables importantes, comme les processus et le contexte.
- **Expliquer** si l'intervention a causé les impacts observés.
- **Synthétiser** les données probantes pour en tirer un jugement évaluatif général.
- **Rendre compte des résultats et en appuyer l'utilisation.**

Cette note d'orientation présente chacun de ces aspects et donne des informations sur une gamme de méthodes qui peuvent être utilisées pour ces aspects. On trouvera des liens vers des sources d'informations supplémentaires. La Note d'orientation 3 étudie comment une approche utilisant des méthodes mixtes, combinant des données quantitatives et qualitatives de manière complémentaire, peut être à la fois mesure/description et explication.

² D'autres ressources sur les méthodes participatives : Marisol Estrella et al. (éds), *Learning from Change : Issues and Experiences in Participatory Monitoring and Evaluation* (Brighton : Institute of Development Studies, 2000), <http://www.idrc.ca/EN/Resources/Publications/Pages/IDRCBookDetails.aspx?PublicationID=348> ; Andrew Catley et al., « Participatory Impact Assessment, » Feinstein International Center, Tufts University : octobre 2008, <http://sites.tufts.edu/feinstein/2008/participatory-impact-assessment> ; Robert Chambers, « Who Counts? The Quiet Revolution of Participation and Numbers, » Document de travail No. 296 (décembre 2007), Brighton : Institute of Development Studies, <http://www.ids.ac.uk/files/Wp296.pdf>.

6. Déterminer les valeurs pour une évaluation d'impact

La première étape est de préciser les valeurs qui sous-tendent l'évaluation. Une évaluation d'impact tire des conclusions sur le niveau de réussite (ou d'échec) d'une intervention et il est donc important de préciser ce que serait la réussite en termes de :

- **Réaliser des impacts souhaitables et éviter (ou au moins minimiser) les impacts négatifs.** Par exemple, la réussite d'un projet de développement de route sera-t-il jugé en termes d'augmentation de l'accès au marché ou d'un meilleur accès aux maternités ? Quel niveau de perte d'habitat et de biodiversité serait considéré comme un coût raisonnable pour la route ? à partir de quel niveau le compromis ne sera-t-il plus acceptable ?
- **Réaliser une distribution souhaitable des bénéfices.** Par exemple, devrait-on juger de la réussite en termes d'effets moyens sur l'éducation, d'améliorations pour les plus désavantagés ou du fait que les membres d'un groupe vulnérable ou désavantagé (comme les petites filles) sont amenés au même niveau que leurs homologues plus avantagés ?

Les objectifs fixés formellement (comme les Objectifs du millénaire pour le développement) et les politiques organisationnelles sont un point de départ intéressant pour préciser les valeurs mais ne suffisent généralement pas à eux seuls. Différentes parties prenantes peuvent avoir des points de vue différents sur les valeurs qui devraient être utilisées dans une évaluation.

Quelques méthodes pour préciser les valeurs pour une évaluation d'impact :

Méthodes qui aident les gens à exprimer des valeurs tacites

Exploration positive – des parties prenantes clés (dont les employés du programme) se souviennent de moments où le programme fonctionnait particulièrement bien puis identifient les valeurs qu'il exemplifiait à ces moments-là.³

Enquêtes dans la communauté – des personnes de la communauté choisissent ou notent les questions qu'ils pensent les plus importantes à traiter.

Changement le plus significatif – un processus structuré pour générer et choisir des histoires de changement qui montrent ce que différentes personnes et différents groupes considèrent comme les résultats ou les impacts les plus importants.⁴

Méthodes qui aident à négocier entre différents systèmes de valeurs

Delphi – processus qui utilise une série d'interactions écrites, sans contact face à face, où les parties prenantes clés donnent leur opinion sur ce qui leur semble important puis réagissent aux résultats combinés.⁵

³ Lecture complémentaire : Hallie Preskill, « Using Appreciative Inquiry in Evaluation Practice, » <http://www.eval.org/summerinstitute07/handouts/sio7.preskillf.pdf>.

⁴ Lecture complémentaire : Most Significant Change (MSC), <http://mande.co.uk/special-issues/most-significant-change-msc/>.

⁵ Lecture complémentaire : Delphi Method, http://www.unido.org/fileadmin/import/16959_DelphiMethod.pdf; The Delphi Method: Techniques and Applications, <http://is.njit.edu/pubs/delphibook/>; Delphi Survey – Europa, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/downloads/sb2_delphi_survey.doc.

Vote avec des gommettes – dans une réunion en face à face, les personnes « votent » (avec des gommettes) pour des options.⁶

7. Élaborer une théorie ou un modèle de la manière dont l'intervention est censée fonctionner

Il est souvent utile de baser une évaluation d'impact sur une théorie ou un modèle de la manière dont l'évaluation devra produire les impacts prévus. On peut appeler cela une théorie de programme, une théorie du changement, une chaîne des résultats ou un modèle logique. Il vaut mieux élaborer la théorie du changement au cours de la planification de l'intervention puis la revoir et la réviser si nécessaire quand on prépare l'évaluation d'impact. Si cela n'a pas été fait au moment où l'intervention démarre, il est possible de s'accorder rétroactivement sur une théorie du changement.

Selon le moment où la théorie du changement est élaborée, elle peut être basée sur une combinaison de sources : documents officiels et objectifs fixés ; recherche sur des interventions similaires ; observations de l'intervention ou d'interventions similaires ; demander à différentes parties prenantes (comme les planificateurs, le personnel et les bénéficiaires prévus) comment elles pensent que cela fonctionne (ou devrait fonctionner).

Il peut y avoir plusieurs théories du changement : différentes théories montrant comment l'intervention fonctionne à différentes étapes, dans différentes situations (en prenant en compte les effets des influences externes) et pour des impacts différents, et différentes théories qui sont

élaborées sur la durée, au fur et à mesure que l'on comprend mieux.

Les théories du changement peuvent améliorer l'évaluation d'impact en permettant :

- D'identifier les effets ou les impacts immédiats qui peuvent être observés durant la période de l'évaluation et qui sont des pré-curseurs des impacts à plus long terme que l'intervention devrait produire.
- De déterminer, quand une intervention n'a pas réussi, à quel point dans le processus elle a cessé de fonctionner ou a été défaillante.
- De faire la distinction entre un échec dû à la *mise en œuvre* (quand l'impact n'a pas été réalisé parce que l'intervention n'a pas été mise en œuvre correctement) et un échec dû à la *théorie* (quand l'intervention ne mène pas à l'impact souhaité, même si elle est mise en œuvre correctement).
- De déterminer quels sont les aspects de l'intervention qui la font fonctionner et qui sont donc cruciaux et doivent être continués quand une intervention est adaptée pour d'autres situations.
- D'identifier les variables comportementales et contextuelles importantes qui devraient être prises en compte dans la collecte, l'analyse et le rapportage des variables pour comprendre les variations d'impact.
- De fournir un cadre conceptuel pour combiner les données probantes diverses correspondant à un programme composé de nombreuses interventions diverses.

Quelques méthodes pour représenter une théorie du changement :

Approche du cadre logique (logframe) – le format classique utilisé dans de nombreuses

⁶ Lecture complémentaire : NRCOI Quick Tip, <http://centerfor-childwelfare.fmhi.usf.edu/kb/bpam/NRCOI-StickyDotTechnique.pdf>.

organisations qui utilise un tableau 4x4. Les quatre lignes sont *activités, produits, objectif et but* et les quatre colonnes sont *une description narrative, des indicateurs objectivement vérifiables (IOV), les moyens de vérification (MV) et les hypothèses.*⁷

Chaine des résultats – l'intervention est représentée sous forme d'une série de cadres en séquence : intrants, activités, produits, résultats à court terme, résultats à plus long terme et impacts.⁸

Chaine des effets /hiérarchie des effets/ chaine du changement – la théorie est représentée sous forme d'une série d'effets intermédiaires menant aux impacts prévus. Ce format attire l'attention sur la manière dont le changement se produit et il est utile pour représenter des programmes où différentes activités ont lieu tout le long de la chaine causale et pas seulement au début.⁹

Cartographie des effets – elle identifie les « partenaires limitrophes » – des organisations ou des groupes dont les actions ne sont pas sous le contrôle de l'intervention et qui sont essentiels pour l'impact à réaliser – puis voir ce que ces partenaires doivent faire et comment l'intervention peut tenter de les influencer.¹⁰

⁷ Lectures complémentaires : The Logical Framework Approach, http://www.ausaid.gov.au/ausguide/pdf/ausguideline3_3.pdf; <http://www.gdrc.org/ngo/logical-fa.pdf>; Logical Framework Analysis, <http://www.gdrc.org/ngo/logical-fa.pdf>; Beyond Lograme : Critique, Variations and Alternatives : <http://www.fasid.or.jp/shuppan/hokokusho/pdf/h21-3.pdf>.

⁸ Lectures complémentaires : Results Chain: Enhancing Program Performance with Logic Models Guide, <http://www.uwex.edu/ces/Imcourse/>; W.K. Kellogg Foundation Logic Model Development Guide, <http://www.evaluationtoolsforracialequity.org/evaluation/tool/doc/pub3669.pdf>.

⁹ Lectures complémentaires : Theories of Change and Logic Models: Telling Them Apart, http://www.evaluationtoolsforracialequity.org/evaluation/resource/doc/TOCs_and_Logic_Models_forAEA.ppt.

¹⁰ Lectures complémentaires : Outcome Mapping, <http://publicwebsite.idrc.ca/EN/Programs/Evaluation/Pages/IDRCBook-Details.aspx?PublicationID=121>; Outcome Mapping: ILAC Brief 7, http://www.outcomemapping.ca/download.php?file=/resource/files/csette_en_ILAC_Brief07_mapping.pdf.

On peut trouver d'autres ressources utiles pour élaborer une théorie du changement dans « Developing a Logic Model or Theory of Change. »¹¹

Une théorie du changement peut aussi être utilisée pour gérer des impacts négatifs potentiels ou pour planifier une évaluation d'impact qui les mesure.

Par exemple, un programme devant améliorer la productivité agricole en encourageant les paysans à utiliser des engrais dans leurs champs peut entraîner une augmentation des ruissellements de phosphates et une pollution des cours d'eau. Une évaluation équilibrée de l'impact étudiera cet impact possible en plus de l'impact d'amélioration de la productivité qui était prévu. On peut construire une théorie du changement pour examiner la manière dont une intervention peut causer des impacts négatifs. Cela pourra servir à adapter l'intervention pour minimiser ou éviter ces impacts négatifs, élaborer des indicateurs d'alerte précoce pour le suivi et s'assurer que ceux-ci sont inclus dans le plan d'évaluation d'impact.

8. Mesurer ou décrire les impacts (et autres variables importantes)

Une évaluation d'impact a besoin de données probantes crédibles, pas seulement au sujet des impacts. Il faut aussi de bonnes informations montrant dans quelle mesure une intervention a été bien mise en œuvre pour pouvoir faire la distinction entre un échec dû à la mise en œuvre ou un échec dû à la théorie. Il faut aussi des informations sur le contexte pour comprendre si une intervention ne fonctionne que dans des situations particulières.

¹¹ The Community Toolbox, "Developing a Logic Model or Theory of Change," http://ctb.ku.edu/en/tablecontents/sub_section_main_1877.aspx.

Il est utile d'identifier des données sur les impacts, la mise en œuvre et le contexte qui seraient déjà disponibles auprès de sources existantes, comme des statistiques officielles, la documentation du programme, un système d'information géographique (SIG) et des projets antérieurs d'évaluation et de recherche. On peut collecter des données supplémentaires pour combler les lacunes ou améliorer la qualité des données existantes en utilisant des méthodes telles que des entretiens (individuels ou de groupe ; structurés, semi structurés ou non structurés), des questionnaires (comme des questionnaires sur internet et une collecte de données par téléphone portable), l'observation (structurée, semi structurée ou non structurée) et des mesures directes (par exemple de la qualité de l'eau par rapport à une norme internationale).

La description des impacts ne devrait pas donner seulement la moyenne mais aussi montrer les variations dans les résultats et en particulier rendre compte des tendances. Howard White explique dans son article de 2010 qu'il est important d'observer l'hétérogénéité :

« Une étude qui présente une seule estimation d'impact (l'effet moyen du traitement)¹² risque d'être moins utile pour les décideurs politiques qu'une étude qui étudierait dans quel contexte les interventions sont le plus efficaces, quels groupes cibles bénéficient le plus et quelles situations environnementales sont utiles ou préjudiciables pour atteindre l'impact. On pourrait donc montrer qu'une intervention éducative, comme des tableaux de conférence, fonctionne, mais seulement si les enseignants ont eux-mêmes un certain

niveau d'éducation ou seulement si les écoles sont déjà bien équipées en matériels de lecture ou si les parents des élèves sont eux-mêmes éduqués. »¹³

Voici quelques sources pour trouver des mesures et des indicateurs pour des secteurs particuliers :

Catalogue de questionnaires d'enquêtes : International Household Survey

Network – plus de 2000 questionnaires qui peuvent être recherchés par pays, par date ou par type d'enquête.¹⁴

Gouvernance démocratique – Centre de Gouvernance du PNUD, Oslo.¹⁵

Indice de pauvreté humaine – trois indicateurs sur la survie, la connaissance et l'impossibilité d'accéder à ce que procure l'économie, PNUD.¹⁶

Objectifs du millénaire pour le développement – 48 indicateurs techniques et 18 cibles pour les 8 objectifs.¹⁷

Développement durable – 130 indicateurs sur les aspects sociaux, économiques, environnementaux et institutionnels du développement durable, Commission des Nations Unies sur le développement durable.¹⁸

13 Howard White, « A Contribution to Current Debates in Impact Evaluation, » *Evaluation* (avril 2010, vol. 16 No 2) : 160.

14 http://www.ihsn.org/home/index.php?q=country_questionnaires

15 http://www.undp.org/oslocentre/flagship/democratic_governance_assessments.html

16 <http://hdr.undp.org/en/statistics/indices/hpi/>

17 Projet du millénaire : <http://www.unmillenniumproject.org/goals/gti.htm>

18 http://www.un.org/esa/sustdev/natlinfo/indicators/methodology_sheets.pdf

12 L'effet moyen du traitement est une estimation de la différence moyenne que cause une intervention. Par exemple, les élèves participant au programme sont restés à l'école une moyenne de 2,5 ans de plus (par rapport au groupe de contrôle).

Indicateurs mondiaux du développement – La Banque mondiale a des données sur plus de 200 pays et sur plus de 1000 indicateurs.¹⁹

La note d'orientation 3 donne plus de détails sur des méthodes particulières pour mesurer ou décrire les impacts, sur l'utilisation de méthodes mixtes (données quantitatives et qualitatives utilisées de manière complémentaires) et sur des manières de résoudre les difficultés de mesure ou de description des impacts.

9. Expliquer dans quelle mesure les résultats observés ont été produits par l'intervention

Une des caractéristiques importantes d'une évaluation d'impact est qu'elle ne se contente pas de collecter des preuves que les impacts se sont bien produits mais qu'elle essaie de comprendre le rôle de l'intervention dans la production de ces impacts. Il est rare qu'une intervention soit la seule cause des changements. Habituellement, une intervention se combine avec d'autres programmes, un contexte favorable et d'autres facteurs. Souvent, un groupe collabore pour produire un impact conjoint, par exemple quand des ONG internationales travaillent en partenariat avec des gouvernements locaux et des communautés. L'attribution des causes n'est donc généralement pas l'attribution totale (c'est à dire que l'intervention était la cause unique) mais une attribution partielle ou une analyse de la contribution de l'intervention. On parle parfois de « contributions plausibles. »

Par exemple, dans la recherche agricole, les impacts en termes d'amélioration de la productivité peuvent être dus à une longue chaîne de recherche de base et de recherche appliquée, d'élaboration de produit et de communication.

¹⁹ <http://data.worldbank.org/data-catalog/world-development-indicators>

Un investissement dans n'importe lequel de ces éléments pourrait raisonnablement affirmer qu'il a été essentiel pour la production des impacts, mais il n'aurait pas pu le faire sans les autres interventions. Autrement dit, il s'agirait d'une intervention nécessaire mais qui n'aurait pas été suffisante pour causer l'impact à elle seule.

Il peut être utile d'étudier l'attribution des causes ou la contribution plausible par rapport à trois éléments. Le point de départ est le factuel : comparer les résultats réels à ceux qui étaient attendus si la théorie du changement était vraie. Quand et où les impacts se sont-ils produits et pour qui ? Ces résultats correspondent-ils à la théorie que l'intervention a causé ces résultats ou y a contribué ? Le second élément est contre-factuel : une estimation de ce qui se serait passé en l'absence d'intervention. Le troisième élément consiste à enquêter pour exclure d'autres explications. Dans certains cas, il sera possible d'inclure les trois éléments dans une évaluation d'impact. Dans des situations complexes, il ne sera peut-être pas possible d'estimer un élément contrefactuel et l'analyse causale devra être basée sur les autres éléments.

Méthodes possibles pour examiner le factuel (dans quelle mesure les résultats réels correspondent à ce qui était prévu) :

Études de cas comparatives – l'intervention n'a-t-elle produits des résultats que dans des cas où d'autres éléments nécessaires étaient en place ?

Dose-réponse – les résultats étaient-ils meilleurs pour les participants qui ont reçu plus de l'intervention (par exemple qui ont participé à plus d'ateliers ou ont reçu plus d'appui) ?

Attribution des bénéficiaires/ des experts – les participants ou les informateurs clés pensent-ils que l'intervention a fait une différence et pourraient-ils expliquer de façon plausible pourquoi ?

Prédictions – les participants ou les sites dont on pensait qu'ils bénéficieraient des meilleurs impacts (grâce à la qualité de la mise en œuvre et/ ou à un contexte favorable) ont-ils vraiment connu ces impacts ? Comment est-il possible d'expliquer les anomalies ?

Temporalité – les impacts se sont-ils produits à un moment correspondant à la théorie du changement – pas avant la mise en œuvre de l'intervention ?

Méthodes possibles pour examiner le scénario contrefactuel (estimation de ce qui se serait passé en l'absence d'intervention) :

Différence dans la différence – La différence avant/après pour le groupe bénéficiant de l'intervention (là où il n'a pas été choisi de manière aléatoire) est comparée à la différence avant/après pour ceux qui n'ont pas bénéficié de l'intervention.²⁰

Contrefactuel construit logiquement – Dans certains cas, il est crédible d'utiliser les données de base comme estimation du scénario contrefactuel. Par exemple, quand une pompe à eau a été installée, il

peut être raisonnable de mesurer l'impact en comparant le temps passé à aller chercher de l'eau à une pompe éloignée avant et après l'intervention, parce qu'il n'y a pas de raison crédible expliquant pourquoi le temps passé aurait diminué sans l'intervention (White, 2007). La reconstitution de processus peut appuyer cette analyse à chaque étape de la théorie du changement.²¹

Comparaisons avec des groupes témoins – Les participants (individus, organisations ou communautés) sont comparés à des non participants pour des variables dont on pense qu'elles sont intéressantes. Il peut être difficile de faire une comparaison adéquate sur tous les critères intéressants.²²

Données de bases multiples ou tournantes – La mise en œuvre d'une intervention est échelonnée dans le temps et dans les populations de l'intervention. L'analyse cherche la répétition dans chaque communauté d'une tendance de changement dans le résultat mesuré après la mise en œuvre de l'intervention ainsi que l'absence de fluctuation substantielle dans les données à d'autres points dans le temps. Ceci est de plus en plus utilisé pour les interventions de santé au niveau de toute une population.²³

²⁰ Lectures complémentaires : Difference-in-differences, http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/counterfactual_impact_evaluation/difference-in-differences/difference-in-differences_en.htm.

²¹ Lectures complémentaires : Process Tracing in Case Study Research, <http://users.polisci.wisc.edu/kritzer/teaching/ps816/ProcessTracing.htm>.

²² Lectures complémentaires : Techniques for Improving Constructed Matched Comparison Group Impact/Outcome Evaluation Designs, <http://knol.google.com/k/techniques-for-improving-constructed-matched-comparison-group-impact-outcome>.

²³ Lectures complémentaires : The Multiple Baseline Design for Evaluating Population-Based Research, <http://rds.epi-ucsf.org/ticr/syllabus/courses/23/2011/06/02/Lecture/readings/Hawkins.pdf>.

Taux de propension – Cette technique crée statistiquement des groupes comparables sur la base d'une analyse des facteurs qui influent sur la propension des gens à participer au programme. Cela est particulièrement utile quand la participation est volontaire (par exemple regarder un programme télévisé qui présente des messages de promotion de la santé).²⁴

Essai contrôlé aléatoire (ECA) – Les participants (ou les communautés, ou les ménages) potentiels sont choisis de façons aléatoire pour bénéficier de l'intervention ou pour être dans un groupe de contrôle (soit pas d'intervention, soit l'intervention habituelle) et les résultats moyens des différents groupes sont comparés.²⁵

Discontinuité de la régression – Quand une intervention n'est disponible que pour des participants au dessus ou en dessous d'un seuil particulier (par exemple de revenu), cette approche compare les résultats des individus juste en dessous du seuil avec ceux des personnes juste au dessus.²⁶

Scénario contrefactuel créé statistiquement – Un modèle statistique, comme une analyse de régression, est utilisé pour établir une estimation de ce qui se serait

passé en l'absence d'intervention. Ceci peut être utilisé quand l'intervention est déjà à une certaine échelle, par exemple une évaluation de l'impact de la privatisation des services nationaux d'approvisionnement en eau.

Il peut être difficile, dans la pratique, d'élaborer un élément contrefactuel crédible. Il est souvent difficile de faire correspondre des individus ou des communautés sur des variables qui font vraiment une différence. Les essais contrôlés aléatoires peuvent créer aléatoirement des groupes non équivalents. D'autres méthodes dépendent d'hypothèses variées qui ne seront peut-être pas réalisées. Dans des situations de changement rapide et imprévisible, il ne sera peut-être pas possible de construire un scénario contrefactuel crédible. Il sera peut-être possible de montrer de façon convaincante et empirique qu'une intervention a produit certains impacts, mais pas d'être sûr de ce qui se serait passé si l'intervention n'avait pas été mise en œuvre. Par exemple, il sera possible de montrer que le développement d'infrastructures communautaires pour l'élevage de poisson pour la consommation ou la vente est dû directement à un projet local, sans pouvoir affirmer avec certitude que cela ne se serait pas passé en l'absence du projet (peut-être grâce à un autre projet mis en œuvre par une autre organisation). Ce sur quoi une évaluation d'impact peut se concentrer, ce sont les deux éléments de l'analyse causale – les alternatives *factuelle* et *d'exclusion*.

Le troisième élément pour comprendre les liens causaux est d'étudier et d'exclure des explications alternatives. Des impacts apparents (ou leur absence) peuvent refléter des problèmes méthodologiques, comme des biais dans la sélection (quand les participants sont systématiquement différents des non participants) et des effets de contamination (quand les non participants bénéficient aussi de l'intervention,

²⁴ Lectures complémentaires : Propensity Scores: What Do They Do, How Should I Use Them, and Why Should I Care?, <http://www.chrp.org/love/ASACleveland2003Propensity.pdf>; A Practical Guide to Propensity Score Models, <http://www.academyhealth.org/files/2009/monday/HebertP.pdf>.

²⁵ Lectures complémentaires : Using Randomization in Development Economics Research : A Toolkit, <http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf>.

²⁶ Lectures complémentaires : Impact Evaluation: Regression Discontinuity, <http://www.urban.org/toolkit/data-methods/regression.cfm>.

ce qui diminue la différence d'impact entre eux et les participants). Ils peuvent refléter l'influence d'autres facteurs, comme d'autres interventions ou des mouvements de population entre des zones où une intervention est mise en œuvre et d'autres où il n'y en a pas.

Méthodes possibles pour identifier et exclure d'autres explications possibles :

Méthodologie générale d'élimination – des explications alternatives possibles sont identifiées puis étudiées pour voir si elles peuvent être exclues.²⁷

Chercher des preuves qui infirment/ Faire le suivi des exceptions²⁸

Une évaluation de l'impact de la législation rendant le port du casque obligatoire pour les cyclistes a montré qu'il y avait eu une diminution significative du nombre de blessures à la tête parmi ceux-ci. Cela concordait avec la théorie du changement, mais une explication alternative était que le nombre total de blessures avait diminué du fait de l'augmentation de la construction de pistes cyclables pendant la même période. L'étude des blessures graves a montré que, alors que le nombre de blessures à la tête avait diminué durant cette période, le nombre d'autres types de blessures était resté stable, ce qui confirmait la théorie que c'étaient bien les casques qui avait causé le changement. (Walter et al., 2011)

²⁷ Lectures complémentaires : Can We Infer Causation from Cross-Sectional Data ?, http://www7.national-academies.org/BOTA/School-Level%20Data_Michael%20Scriven-Paper.pdf

²⁸ Lectures complémentaires : Matthew B. Miles et A. Michael Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*, 2ème édition (Thousand Oaks, Californie : Sage Publications, 1994).

Quelques approches qui combinent ces différents éléments d'explication :

Méthodologie MLE (Multiple lines and levels of evidence Lignes et niveaux de données probantes multiples) – une large gamme de données probantes venant de différentes sources est étudiée par un panel d'experts reconnus d'un ensemble de disciplines. Le panel identifie la cohérence avec la théorie du changement tout en identifiant et en expliquant les exceptions. Pour trouver une relation causale entre une intervention et les impacts observés, la méthode MLE étudie la force des données probantes, leur cohérence, leur spécificité, leur temporalité, leur cohérence avec d'autres éléments probants acceptés, leur plausibilité et leur analogie avec des interventions similaires.²⁹

Analyse des contributions – approche systématique qui comprend l'élaboration d'une théorie du changement, la cartographie des données existantes, l'identification des défis à la théorie – comme des manques de données probantes et des liens causaux contestés – et la collecte itérative de données probantes supplémentaires pour y répondre. La Note d'orientation 2 donne des informations complémentaires sur l'analyse des contributions.³⁰

²⁹ Lectures complémentaires : « Patricia Rogers, Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation », *Journal of Development Effectiveness*, 1 (2009) : 217-226. Document de travail disponible sur : http://www.3ieimpact.org/admin/pdfs_papers/50.pdf.

³⁰ Lectures complémentaires : Contribution Analysis : ILAC Guide Brief 16, http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf; Contribution Analysis, <http://www.scotland.gov.uk/Resource/Doc/175356/0116687.pdf>.

Rapports collaboratifs sur les effets –

cette nouvelle approche combine l'analyse des contributions et la méthode MLLE. Elle cartographie les données existantes par rapport à la théorie du changement et comble les manques importants de données probantes en faisant une collecte ciblée de données supplémentaires. Puis elle utilise une combinaison de revue par des experts et de consultation dans la communauté pour vérifier si les données sur les impacts qui se sont produits sont crédibles et dans quelle mesure ceux-ci peuvent réalistement être attribués à l'intervention.³¹

L'évaluation de l'impact d'un programme de développement exécutif intergouvernemental n'a pas pu utiliser un groupe de contrôle aléatoire parce qu'il était impossible de sélectionner aléatoirement les personnes qui seraient dans le groupe de contrôle, ou même qui participeraient au programme. L'évaluation ne pouvait pas non plus utiliser un groupe de comparaison parce que la nature du programme était telle que ceux qui étaient acceptés dans le programme étaient systématiquement différents de ceux qui n'étaient pas acceptés. L'évaluation a donc utilisé d'autres stratégies d'explication causale, comme l'attribution par les bénéficiaires, la temporalité et la spécificité (les changements étaient des les zones spécifiques où travaillaient le programme). (Davidson, 2006)

³¹ Lectures complémentaires : Collaborative Outcomes Reporting Technique, <http://www.aes.asn.au/regions/vic/2010/COR%20AES%202010.pdf>.

10. Synthétiser les données probantes

On évalue rarement toute une intervention uniquement sur une seule mesure de performance. Il faut généralement synthétiser les données probantes sur la performance dans des dimensions différentes.

Pour cela, une manière habituelle de faire est d'élaborer un barème avec des coefficients où un pourcentage de la note générale de performance est basé sur chaque dimension. Cependant, un barème numérique avec coefficients pose souvent des problèmes, comme des coefficients arbitraires et un manque d'attention à des éléments essentiels.³²

Une autre possibilité est d'élaborer un barème global (ou une rubrique) d'évaluation sur lequel on se met d'accord avec les utilisateurs prévus pour qu'il puisse servir à synthétiser de manière transparente les données probantes. Ce barème comprend une appréciation pour chaque point (par exemple, « pas réussi », « assez réussi », « très réussi ») et une description de chacune de ces appréciations.³³

11. Rendre compte des constatations et en encourager l'utilisation

Il faudrait décider du format du rapport d'évaluation dès la planification de l'évaluation d'impact. Certaines organisations ont des formats standard pour les rapports, avec des longueurs précises à respecter. Dans les autres cas, il est important de se mettre d'accord, bien avant la rédaction du

³² Lectures complémentaires : The Synthesis Problem ; Issues and Methods in the Combination of Evaluation Results Into Overall Evaluative Conclusions, <http://realevaluation.com/pres/SynthHNL.pdf>.

³³ Lectures complémentaires : The Rubric Revolution, <http://kinnect.co.nz/wp-content/uploads/2011/09/AES-2011-Rubric-Revolution-Davidson-Wehipeihana-McKegg-xx.pdf>.

rapport, sur un « squelette de rapport » comprenant les titres et sous-titres.

Les rapports d'évaluation d'impact sont plus accessibles quand ils sont organisés autour des questions clés de l'évaluation plutôt que s'ils rendent compte séparément des données sur les différents éléments de la collecte des données.³⁴

La qualité des rapports d'évaluation peut être améliorée en impliquant de façon adéquate les parties prenantes. Même quand l'évaluation est faite par un évaluateur externe indépendant, on peut impliquer les parties prenantes en leur demandant de réagir de manière formelle aux constatations et de faire des commentaires sur les données et la manière dont elles ont été interprétées.

Quand les rapports d'évaluation comprennent des recommandations, celles-ci doivent être justifiées par des données probantes tirés des constatations de l'évaluation et des données probantes sur la faisabilité et l'à-propos des recommandations. En impliquant les parties prenantes concernées dans l'élaboration des recommandations, on n'améliore pas seulement la faisabilité des recommandations. Cela peut aussi conduire les parties prenantes à se les approprier et à s'engager à les mettre en pratique.

L'utilisation de rapports d'évaluation d'impact peut être améliorée si l'on utilise des formats créatifs pour rendre compte, des présentations orales, des possibilités de discuter avec d'autres des implications des évaluations d'impact, et si

³⁴ E. Jane Davidson, « Improving Evaluation Questions and Answers : Getting Actionable Answers for Real-World Decision Makers » (session de démonstration de la Conférence 2009 de l'Association américaine d'évaluation 18 novembre 2009, Orlando, Floride). <http://comm.eval.org/resources/viewdocument/?DocumentKey=e5bac388-f1e6-45ab-9e78-10e60cea0666>.

l'on s'assure que les rapports restent accessibles aux utilisateurs potentiels.

La Note d'orientation 4 parle de la nécessité de bien communiquer les constatations aux publics prévus.

12. Quand faudrait-il faire une évaluation d'impact ?

Les évaluations d'impact devraient être faites quand il y a un vrai besoin et une vraie intention d'utiliser les constatations. Si l'on exigeait une évaluation d'impact pour toutes les interventions, les évaluateurs risqueraient, soit d'avoir besoin de trop de ressources, soit de les disperser tellement que les évaluations seraient superficielles. Une stratégie plus efficace est de diriger les ressources destinées aux évaluations d'impact vers les interventions où elles ont plus de chances d'être utiles :

- Des interventions novatrices et des programmes pilotes qui, s'ils s'avèrent réussis, peuvent être étendus ou répliqués.
- Des interventions dont on ne comprend pas bien l'impact et où on a besoin de meilleures informations pour décider s'il faut continuer à les financer ou s'il vaut mieux utiliser leur financement pour d'autres interventions.
- Des évaluations périodiques de l'impact d'un portefeuille d'interventions dans un secteur ou une région pour guider la politique, la conception d'interventions futures et les décisions de financement.
- Des interventions avec un profil plus risqué, comme un investissement important (actuellement et à l'avenir), un fort potentiel d'impacts négatifs significatifs ou des questions politiques sensibles.

- Des interventions où les parties prenantes ont besoin de mieux comprendre les contributions et les perspectives les unes des autres.

Le calendrier d'une évaluation d'impact est important. Si elle est faite trop tôt, il risque de ne pas y avoir suffisamment de données probantes montrant que des impacts se sont produits ou sont maintenus. Si elle est faite trop tard, il peut être difficile de faire un suivi avec les participants et trop tard pour influencer sur des décisions sur la direction de l'intervention à l'avenir. Dans tous les cas, il vaut mieux prévoir l'évaluation d'impact si possible dès le début de l'intervention. Ceci permet de collecter des données probantes tout au long de l'intervention, par exemple des données de base, et cela donne l'option d'utiliser des méthodes comme des essais contrôlés aléatoires qui demandent la création d'un groupe de contrôle choisi de manière aléatoire dès le début de la mise en œuvre.

13. Qu'est-ce qui est nécessaire pour une évaluation d'impact de qualité ?

Il peut être utile de réfléchir à la qualité de l'évaluation par rapport à cinq impératifs concurrents : *utilité, précision, éthique, pratique et responsabilité*.³⁵ Ces cinq normes sont souvent en contradiction. Par exemple, une évaluation d'impact plus complète qui sera plus précise ne sera peut-être pas pratique du point de vue des ressources disponibles, sera peut-être trop indiscreète dans les données collectées ou pourra prendre trop longtemps à finir pour pouvoir fournir des informations sur lesquelles baser des décisions clés sur l'avenir de l'intervention.

³⁵ Joint Committee Standards for Educational Evaluation <http://www.jcsee.org/program-evaluation-standards/program-evaluation-standards-statements>. Ces normes ont été à l'origine élaborées pour l'évaluation de l'éducation mais sont aujourd'hui souvent utilisées plus largement.

Utilité – une bonne évaluation d'impact est utile. L'utilité probable d'une évaluation peut être améliorée en prévoyant dès le début la manière dont elle sera utilisée, en planifiant les liens avec les processus de prise de décision et le calendrier de l'organisation, en sachant pourquoi elle est faite et qui l'utilisera, en impliquant les parties prenantes clés dans le processus puis en choisissant des systèmes et des méthodes pour réaliser cet objectif.

Précision – une bonne évaluation d'impact est rigoureuse. Elle fait attention à tous les impacts importants, en notant les impacts imprévus. Elle fait attention à la distribution des impacts, en remarquant si seules certaines personnes en bénéficient et qui sont ces personnes. La précision suppose l'utilisation de données probantes appropriées, avec des données quantitatives et qualitatives, une bonne interprétation et une transparence par rapport aux sources de données utilisées et à leurs limites. Parmi les stratégies permettant la précision des évaluations, il y a des systèmes permettant de vérifier la qualité des données à l'endroit de la collecte et durant leur traitement et de vérifier qu'on a rendu compte des constatations de manière juste, complète et claire.

Correction (éthique) – les questions éthiques doivent aussi être résolues – comme la confidentialité et l'anonymat ainsi que les effets négatifs potentiels d'une implication dans l'évaluation. Certaines questions éthiques, comme la nécessité de tenir les promesses touchant à la vie privée et à la confidentialité, sont les mêmes pour tous les types d'évaluation et de recherche. Il y en a d'autres qui sont particulières à l'évaluation d'impact. Il y a parfois des inquiétudes sur l'éthique de l'utilisation d'essais aléatoires contrôlés, car il suppose de refuser une intervention à certaines personnes (le groupe de contrôle). Ceci pose moins de problèmes éthiques quand l'accès à

l'intervention doit de toutes façons être rationné. On peut résoudre ce problème en mettant le groupe de contrôle en attente pour qu'il reçoive l'intervention une fois que l'évaluation de la première phase sera terminée (si celle-ci montre que l'intervention est efficace). Cependant, cette stratégie n'est envisageable que si les impacts (ou des indications crédibles de ceux-ci) sont visibles tôt et que l'intervention est encore pertinente pour le groupe de contrôle au moment où l'évaluation sera terminée.

Il peut aussi y avoir des problèmes éthiques concernant les personnes dont les intérêts sont servis par une évaluation. Dans ses Principes directeurs, l'Association américaine d'évaluation les mentionne sous le terme de « Responsabilités pour le bien-être général et public » :

Les évaluateurs expriment et prennent en compte la diversité des intérêts et des valeurs généraux et publics et doivent donc :

- 1. Inclure les points de vue et les intérêts de toute la gamme des parties prenantes.*
- 2. Prendre en compte, non seulement les opérations immédiates et les effets de l'évaluation mais aussi les grandes hypothèses, les implications et les effets secondaires potentiels.*
- 3. Permettre aux parties prenantes d'avoir accès aux informations de l'évaluation et de les diffuser activement et présenter les résultats de l'évaluation sous des formes compréhensibles qui respectent les personnes et tiennent les promesses de confidentialité.*

4. Maintenir un équilibre entre les besoins et les intérêts des clients et des autres parties prenantes.

5. Prendre en compte l'intérêt et le bien public, en dépassant l'analyse des intérêts de parties prenantes particulières pour considérer le bien-être de la société dans son ensemble.

Pour entreprendre une évaluation d'impact, on a généralement besoin d'une autorisation formelle du comité institutionnel d'éthique. Les demandes d'autorisation doivent suivre le format prescrit et traiter des questions de bienfaisance, de justice et de respect.³⁶

Pratique – les évaluations d'impact doivent être pratiques. Elles doivent prendre en compte les ressources qui sont disponibles (temps, argent, expertise et données existantes) et le moment où on a besoin des résultats de l'évaluation pour aider à prendre des décisions. Un travail en partenariat avec un ou plusieurs professionnels de l'évaluation, organisations de recherche, universités et organisations de la société civile, peut aider à procurer les ressources nécessaires.

Responsabilité – les évaluations doivent expliquer clairement sur quelles données probantes et sur quels critères on a basé les conclusions et reconnaître leurs limites. Il est important d'être transparent à propos des sources de données, et de montrer quelles sources ont été utilisées pour quelles questions d'évaluation. Un processus formel de méta-évaluation – faire évaluer sa propre évaluation – en approuvant un plan d'évaluation puis un

³⁶ Lectures complémentaires : Evaluation Consent and the Institutional Review Board Process, <http://meera.snre.umich.edu/plan-an-evaluation/plonearticlemultipage.2007-10-30.3630902539/evaluation-consent-and-the-institutional-review-board-process>.

rapport d'évaluation – par un expert ou un comité de personnes respectées pour leur intégrité et leur indépendance peut améliorer la responsabilité d'une évaluation d'impact.

14. Difficultés courantes dans l'évaluation d'impact dans le développement

On trouvera ci-dessous la description d'un certain nombre de difficultés courantes pour l'évaluation du développement ainsi que des suggestions pour y faire face.

- **Variations dans la mise en œuvre et l'environnement selon les sites**

Une intervention peut avoir été mise en œuvre de manières tout à fait différentes selon les situations différentes dans les différents bureaux de pays du monde ou dans des zones géographiques différentes à l'intérieur d'un même pays. Il peut être utile de comparer les théories du changement de tous les sites. Déterminer en particulier si les différents sites utilisent la même théorie sur la manière dont le changement a lieu (par ex. en permettant aux populations de mieux connaître leurs droits à des services) mais des théories d'action différentes (par ex. des brochures imprimées au lieu de théâtre communautaire) ou s'ils utilisent des théories du changement tout à fait différentes (par ex. faire mieux connaître aux populations leurs droits à des services sur un site et réduire les obstacles à l'accès aux services – comme les frais d'utilisation – grâce à un plaidoyer sur un autre site).

- **Impacts hétérogènes**

Souvent, des interventions de développement ne fonctionnent bien que pour

certaines personnes et peuvent être inefficaces ou même nuisibles pour d'autres. De plus, le fait qu'une intervention réussisse à atteindre les impacts souhaitables est souvent affecté par la qualité de la mise en œuvre. Il est donc important de ne pas seulement faire des calculs et des rapports sur l'effet général mais de vérifier aussi les effets différentiels. Il faut pour cela collecter si possible des données probantes sur la qualité de la mise en œuvre et des données sur des facteurs contextuels qui peuvent avoir un effet sur l'impact, y compris les caractéristiques des participants et l'environnement de mise en œuvre.

- **Éléments divers**

Un programme peut comporter une large gamme de projets et pourtant, il faut faire une évaluation générale de l'impact de tout le programme. Il peut être utile d'élaborer une théorie générale du changement pour tout le programme, en réunissant ses différents éléments. Parfois, il est possible de faire cela dès la phase de planification, mais, surtout quand les projets ou les éléments ont varié avec le temps, il faudra peut-être le faire rétroactivement.

- **Longues échelles de temps**

Souvent les impacts prévus ne seront visibles qu'après de nombreuses années, mais on a besoin avant cela de données probantes pour prendre des décisions en connaissance de cause (par ex. s'il faut ou non mettre en route une phase suivante ou répliquer le modèle ailleurs). Une théorie du changement peut identifier les effets immédiats qui pourraient être visibles sur la durée de l'évaluation. Dans

certains cas, des résultats de recherche peuvent être utilisés pour trouver des liens plus tardifs et estimer les impacts probables d'après les résultats intermédiaires atteints. Il faudrait aussi prendre en compte la trajectoire prévue du changement – quand les impacts ont des chances d'être évidents.³⁷

- **Influence d'autres programmes et facteurs**

Les impacts des interventions de développement sont fortement influencés par les activités d'autres programmes et d'autres facteurs contextuels qui peuvent aider à la réalisation des impacts ou l'empêcher. Par exemple, des transferts d'argent qui sont subordonnés à l'assiduité des élèves à l'école ne permettront à ceux-ci d'avoir de meilleurs résultats que dans des situations où les écoles offrent un enseignement adéquat. Il est possible d'identifier ces autres programmes et facteurs contextuels dans le cadre de l'élaboration d'une théorie du changement, de collecter des données sur ceux-ci et de chercher des tendances dans les données.

- **Contraintes de ressources**

Les éléments probants existants (sous forme de documents de programme, de données de base ou de statistiques officielles) peuvent être incomplets et il peut n'y avoir que peu de ressources (en termes de financement, de temps des employés ou d'accès à une expertise technique spécialisée) pour collecter les types de données nécessaires pour faire une

évaluation d'impact de qualité. Pour une évaluation particulière, quand il n'existe que peu de données probantes et qu'il n'y a que peu de ressources pour collecter des données supplémentaires, des entretiens avec divers informateurs clés peuvent fournir suffisamment de données et permettre de reconstituer les données de base. Planifier l'évaluation d'impact à l'avance peut limiter les contraintes de ressources en prévoyant suffisamment de ressources aux phases de conception et de budget et/ou en allouant les ressources d'évaluation stratégiquement entre les interventions pour qu'elles soient plus concentrées sur un petit nombre d'évaluations plus complètes d'interventions stratégiquement importantes.

Résumé

Une évaluation d'impact devrait commencer par un plan qui précise ses objectifs prévus, avec l'identification des utilisateurs prévus, les questions clés auxquelles elle doit répondre et comment elle traitera les six aspects d'une évaluation d'impact : préciser quelles sont les valeurs, élaborer une théorie du changement, mesurer ou décrire les variables importantes, expliquer ce qui a produit les impacts, synthétiser les données probantes, rendre compte des résultats et en appuyer l'utilisation. Si ce plan est vérifié (y compris par les utilisateurs prévus), cela rendra plus probable la production d'une évaluation d'impact de grande qualité et qui sera réellement utilisée.

Références et autres ressources utiles

Alton-Lee, A. (2003) « Quality Teaching for Diverse Students in Schooling : Best Evidence Synthesis. » Wellington, Nouvelle Zélande :

³⁷ Lectures complémentaires : Michael Woolcock on The Importance of Time and Trajectories in Understanding Project Effectiveness, <http://blogs.worldbank.org/impactevaluations/guest-post-michael-woolcock-on-the-importance-of-time-and-trajectories-in-understanding-project-effe>.

- Ministère de l'éducationn. <http://www.educationcounts.govt.nz/publications/series/2515/5959>
Un exemple de la manière de synthétiser des données probantes venant de sources diverse,s pour comprendre ce qui fonctionne pour qui.
- Catley A., Burns, J., Abebe, D., Sufi, O.
Participatory Impact Assessment : A Guide for Practitioners. Boston : Tufts University. <http://www.preventionweb.net/english/professional/publications/v.php?id=9679>
- Chambers R. (2007) « Who Counts ? The Quiet Revolution of Participation and Numbers » Document de travail No 296, Brighton : Institute of Development Studies. <http://www.ids.ac.uk/files/Wp296.pdf>
- Davidson, E. J. (2006) « Causal Inference Nuts and Bolts. » Session de démonstration à la Conférence de 2006 de l'American Evaluation Association, Portland, Oregon, nov. 2006 <http://realevaluation.com/pres/causation-an-zea09.pdf>.
- Davidson, E. J. (2009) « Improving Evaluation Questions and Answers : Getting Actionable Answers for Real-World Decision Makers. » Session de démonstration à la Conférence de 2000 de l'American Evaluation Association, Orlando, Floride, nov. 2009. <http://comm.eval.org/resources/viewdocument/?DocumentKey=e5bac388-f1e6-45ab-9e78-10e60ceao666>.
- Funnell S. et Rogers, P. (2011) *Purposeful Program Theory : Effective Use of Theories of Change and Logic Models*. San Francisco : Jossey-Bass/Wiley.
- Guijt, I. (1999) Participatory Monitoring and Evaluation for Natural Resource Management and Research. Socio-economic Methodologies for Natural Resources Research. Chatham, Royaume Uni : Natural Resources Institute. <http://www.nri.org/publications/bpg/bpgo4.pdf>
- Miles, M. et Huberman, M. (1994) *Qualitative Data Analysis : An Expanded Sourcebook* (2ème éd.) Thousand Oaks Californie : Sage Publications. *Présente des stratégies pour vérifier les explications causales, y compris chercher des preuves qui infirment, faire le suivi des exceptions, faire et tester des prédictions.*
- Patton, MQ (2008) « State of the Art in Measuring Development Assistance. » *Explique l'importance de l'interprétation et de la gestion de l'incertitude dans une gestion efficace.*
- Paz R., Dorward A., Douthwaite B. (2006). « Methodological Guide for Evaluation of Pro-Poor Impact of Small-Scale Agricultural Projects. » Centre for Development and Poverty Reduction. Imperial College, Londres. <http://boru.pbworks.com/f/modulosjano7.pdf> *Décrit 22 méthodes et outils qui peuvent être utilisés pour évaluer les impacts directs et indirects de l'adoption d'innovations.*
- Roche, C. (1999) *Impact Assessment for Development Agencies: Learning to Value Change* Oxford : OXFAM, Novib
- Rogers, Patricia J. (2009) « Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Évaluation », *Journal of Development Effectiveness*, 1(3): 217-226.
- Document de travail disponible sur : http://www.3ieimpact.org/admin/pdfs_papers/50.pdf.

Walter, S., Olivier, J., Churches, T. et Grzebieta, R. (2011). « The impact of compulsory cycle helmet legislation on cyclist helmet head injuries in New South Wales, Australia », *Accident Analysis and Prevention*, 43 : 2064-2071.

White, S. et J. Petit (2004) *Participatory Methods and the Measurement of Wellbeing*
Participatory Learning and Action 50, Londres : IIED

www.betterevaluation.org – informations sur les méthodes d'évaluation pour le développement, avec des exemples fournis par des utilisateurs et des commentaires.

www.mymande.org – informations, vidéos et liens pour des informations sur les méthodes d'évaluation.

