

Présentation des méthodes de collecte et d'analyse de données dans l'évaluation d'impact

Greet Peersman

CENTRE DE RECHERCHE INNOCENTI DE L'UNICEF

Le Centre de recherche Innocenti est le bureau de l'UNICEF spécialisé en recherche scientifique. Les objectifs prioritaires du Centre de recherche sont d'améliorer la compréhension internationale des questions liées aux droits des enfants afin de faciliter la pleine application de la Convention relative à ces droits dans le monde entier. Le Centre a pour objectif de mettre en place un cadre intégré pour la gestion de la recherche et des connaissances au sein de l'organisation, afin de soutenir ses programmes et politiques à l'échelle mondiale. Les publications du Centre, en présentant un vaste éventail d'opinions, contribuent au débat international sur les questions liées aux droits des enfants.

Cette publication ne reflète pas les politiques ou points de vue de l'UNICEF sur certains sujets. Les opinions exprimées n'engagent que leurs auteurs et/ou réviseurs et sont publiées afin d'encourager le dialogue sur les méthodes d'évaluation d'impact.

NOTES MÉTHODOLOGIQUES DU CENTRE DE RECHERCHE INNOCENTI

Les notes méthodologiques du Centre d'Investigations d'UNICEF ont le but de partager des pratiques, des méthodes et des dessins ainsi que des recommandations de chercheurs et analystes reconnus. Elles sont dirigées principalement au personnel de l'UNICEF qui conduit des recherches ou qui interprète les résultats et analyses des évaluations réalisées par des chercheurs externes pour les prises de décision en phase de programmation ou à l'appui des politiques ou de la sensibilisation.

Cette note méthodologique a suivi un procès interne de révision par pairs.

Ce texte n'a pas été édité conformément aux standards officiels de publication et UNICEF décline toute responsabilité pour les possibles erreurs.

La reproduction de parties de cette publication est permise à condition que la source soit dûment citée. Pour l'utilisation d'une partie substantielle ou de la totalité de la publication, veuillez bien vous adresser au Département de Communication au suivant courriel florence@unicef.org

Pour consulter ou télécharger ces notes méthodologiques, visitez <http://www.unicef-irc.org/KM/IE/>

Nous conseillons d'utiliser la citation suivante pour toute référence au document présent:

Peersman, G. (2014). Présentation des méthodes de collecte et d'analyse de données dans l'évaluation d'impact *Note méthodologique n°10*, Centre de recherche Innocenti, Florence.

Remerciements : Ce dossier a profité des conseils de beaucoup d'individus. L'auteur et le Centre de Recherche désirent remercier tous ceux qui ont contribué et en particulier :

Pour leur contribution: Simon Hearn, Jessica Sinclair Taylor

Pour leur révision: Nikola Balvin, Claudia Cappa, Yan Mu

© Fonds des Nations Unies pour l'enfance (UNICEF), septembre 2014
Centre de recherche Innocenti de l'UNICEF

Piazza SS. Annunziata, 12
50122 Florencia (Italia)
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

1. BRÈVE DESCRIPTION DE LA COLLECTE ET DE L'ANALYSE DES DONNÉES

Quel que soit le type d'évaluation menée, il est essentiel de bien choisir les méthodes de collecte et d'analyse des données et de les appliquer correctement. La présente note présente brièvement les questions relatives au choix et à l'utilisation de méthodes pour les évaluations d'impact, c'est-à-dire les évaluations qui fournissent des informations sur les effets à long terme (intentionnels ou non) des programmes ou politiques.

Les évaluations d'impact ne doivent pas se cantonner à déterminer l'ampleur des effets (c'est-à-dire l'impact moyen), mais doivent également identifier qui a bénéficié de ces programmes ou politiques et comment. Il convient de préciser dès le début ce qui constitue une « réussite » et la façon dont les données seront analysées et synthétisées pour répondre aux questions clés d'évaluation. La collecte de données doit en effet permettre d'obtenir l'ensemble des données probantes nécessaires pour porter des jugements appropriés sur le programme ou la politique. En d'autres termes, le cadre d'analyse – la méthode pour analyser le « sens » des données en recherchant des tendances de manière systématique et transparente – doit être spécifié au cours de l'étape de planification. Le cadre de travail précise comment l'analyse de données prend en compte les hypothèses formulées dans la théorie du changement du programme relatives à la méthode adoptée afin de produire les résultats escomptés (voir Note N°2, Théorie du changement). Dans une véritable évaluation à méthodes mixtes, ceci inclut le recours à des méthodes d'analyse numérique et textuelle appropriées et à la triangulation de plusieurs sources de données et perspectives afin de maximiser la crédibilité des résultats de l'évaluation.

Points principaux

1. Les méthodes de collecte et d'analyse de données doivent être choisies en fonction de l'évaluation concernée, de ses questions clés d'évaluation et des ressources disponibles.
2. Les évaluations d'impact doivent utiliser autant que possible les données existantes, puis s'appuyer sur de nouvelles données pour combler les lacunes.
3. Les méthodes de collecte et d'analyse de données doivent être sélectionnées de manière à assurer la complémentarité entre les points forts et les faiblesses des unes et des autres.

2. PLANIFIER LA COLLECTE ET L'ANALYSE DES DONNÉES

Commencer avec la planification globale de l'évaluation

Avant de décider des données à recueillir et de la façon de les analyser, il convient de déterminer les buts de l'évaluation (c'est-à-dire ses destinataires et utilisations prévues) et les questions clés d'évaluation (voir Note N°1, Présentation de l'évaluation d'impact). Une évaluation d'impact peut être réalisée en vue d'éclairer les décisions relatives aux changements à apporter à un programme ou à une politique (évaluation formative) ou dans le but de décider de la poursuite, de l'interruption, de la reproduction ou de l'extension ou non d'un programme ou d'une politique (évaluation sommative). Dès que le but de l'évaluation est clairement défini, il est nécessaire de convenir d'un petit nombre de

questions clés d'évaluation de haut niveau (10 au maximum), idéalement avec les parties prenantes concernées ; il se peut que ces questions aient déjà été définies par un système d'évaluation ou un cadre d'évaluation précédemment élaboré. Le fait d'y répondre – peu importe comment elles ont été définies – devrait garantir le respect du but de l'évaluation. Disposer d'un ensemble reconnu de questions clés d'évaluation fournit une orientation sur les données à recueillir et la manière de les analyser et d'en rendre compte dans les résultats de l'évaluation.

Une [théorie du changement](#) correctement élaborée constitue l'un des outils essentiels d'une évaluation d'impact. Elle décrit comment le programme ou la politique est censé(e) fonctionner : elle présente un modèle de causalité qui relie les [intrants](#) et les [activités](#) aux [extrants](#) et aux [effets directs](#) souhaités ainsi qu'aux [impacts](#) (voir Note N°2, Théorie du changement). La théorie du changement doit également tenir compte des résultats (positifs ou négatifs) imprévus. Cet outil est utile non seulement lors de la conception du programme, mais il contribue également à axer l'évaluation d'impact sur les connaissances dont doivent disposer les parties prenantes sur le programme ou la politique afin de prendre des décisions éclairées – en d'autres termes, les questions clés d'évaluation. De bonnes questions d'évaluation ne portent pas uniquement sur les résultats obtenus (questions descriptives), mais également sur la qualité des résultats (appréciation de la valeur du programme ou de la politique). Les évaluations d'impact doivent rassembler des données probantes sur les impacts (p. ex., évolution positive du taux de mortalité des enfants de moins de cinq ans) et examiner comment les impacts prévus ont été atteints, ou pourquoi ils ne l'ont pas été. Il faut pour cela que des données sur le contexte soient disponibles (p. ex., le cadre normatif et juridique qui dans un pays affecte la protection de l'enfant), que les activités de programme ou la mise en œuvre de la politique soient adéquates et de qualité, et qu'il existe divers résultats intermédiaires (p. ex., recours plus important à la vaccination), variables explicatives dans la chaîne de causalité¹.

Utiliser au maximum les données existantes

Il faut commencer la planification de la collecte de données en examinant dans quelle mesure il est possible d'utiliser les données existantes. Concernant les indicateurs, l'évaluation doit s'efforcer de s'appuyer sur différents types [d'indicateurs](#) (intrants, extrants, effets directs, impacts) afin de refléter les résultats clés dans la théorie du changement du programme. Dans l'idéal, les évaluations d'impact doivent utiliser les indicateurs sélectionnés pour suivre la performance tout au long de la période de mise en œuvre du programme, c'est-à-dire les indicateurs clés de performance (ICP). Dans bien des cas, il est également possible d'utiliser les données recueillies par le biais d'enquêtes démographiques standardisées, telles que l'enquête en grappes à indicateurs multiples (MICS), l'enquête démographique et de santé (EDS) ou l'étude sur la mesure des niveaux de vie (LSMS) de l'UNICEF.

Il est particulièrement important de vérifier si des [données de référence](#) sont disponibles pour les indicateurs sélectionnés ainsi que pour les caractéristiques sociodémographiques et autres caractéristiques pertinentes de la population étudiée. Si le modèle d'évaluation prévoit de comparer les changements au fil du temps dans différents groupes, il est alors possible d'utiliser des données de référence pour déterminer l'équivalence des groupes avant le début du programme ou pour « faire correspondre » différents groupes (par exemple dans le cas de [méthodes quasi expérimentales](#) ; voir Note N°8, Modèles et méthodes quasi expérimentaux). Ces données sont également indispensables pour déterminer s'il y a eu une évolution entre deux périodes données et, le cas échéant, son ampleur (c'est-à-dire la [taille de l'effet](#)). En l'absence de données de référence, des données supplémentaires doivent être recueillies afin de reconstituer les données de référence, par exemple grâce au « rappel » (qui consiste à demander aux gens de se remémorer des informations spécifiques

¹ La Note N°1, Présentation de l'évaluation d'impact, porte sur la nécessité d'avoir recours à des approches différentes pour évaluer les politiques plutôt que les programmes.

concernant une expérience ou un événement survenu dans le passé). Il est possible que cette méthode induise un certain biais, qui peut toutefois être considérablement réduite : en étant réaliste quant à ce que les personnes interrogées peuvent se remémorer et ce qu'elles sont moins susceptibles de se rappeler, et en utilisant des outils d'évaluation existants².

Il existe d'autres sources de données, notamment : les statistiques officielles, les données de suivi du programme, les dossiers du programme (qui peuvent inclure une description du programme, une théorie du changement, les comptes rendus de réunions pertinentes, etc.), des documents officiels relatifs à la politique et les plans de mise en œuvre du programme, ainsi que les rapports d'activité. S'il est important d'utiliser au mieux les données existantes par souci d'efficacité, il est nécessaire de disposer de données d'une qualité suffisante pour ne pas compromettre la validité des résultats de l'évaluation (voir ci-après).

Identifier et combler les importantes lacunes en matière de données

Après avoir passé en revue les informations disponibles, il convient de créer une matrice d'évaluation (voir tableau 1) précisant les méthodes de collecte et d'analyse de données utilisées pour répondre à chaque question clé d'évaluation, puis d'identifier et de hiérarchiser les lacunes devant être comblées en recueillant de nouvelles données. Ceci permet de confirmer que la collecte de données prévue (et le regroupement de données existantes) recouvre l'ensemble des questions clés d'évaluation, de déterminer si la triangulation entre les différentes sources de données est suffisante, et de contribuer à la conception d'outils pour la collecte de données (questionnaires, questions pour les entretiens, outils d'extraction des données pour l'examen des documents et outils d'observation par exemple) afin de s'assurer que les informations nécessaires sont recueillies.

Tableau 1. Matrice d'évaluation : faire correspondre la collecte de données aux questions clés d'évaluation

Exemples de questions clés d'évaluation	Enquête sur les participants au programme	Entretiens avec les informateurs clés	Dossiers du projet	Observation de la mise en œuvre du programme
Question 1 Quelle a été la qualité de la mise en œuvre ?		✓	✓	✓
Question 2 Dans quelle mesure les objectifs du programme ont-ils été atteints ?	✓	✓	✓	
Question 3 Quels ont été les autres impacts du programme ?	✓	✓		
Question 4 Comment le programme peut-il être amélioré ?		✓		✓

² White, Howard, « A contribution to current debates in impact evaluation », *Evaluation*, volume 16, n° 2, 2010, p. 153 à 164.

Il existe de nombreuses méthodes différentes pour recueillir des données. Le tableau 2 fournit des exemples de sources de données possibles³ (existantes et nouvelles).

Tableau 2. Options de collecte (données primaires) et de regroupement (données secondaires) de données

Option	Sources et méthodes	Exemples
Recherche de documents et de données existants	<ul style="list-style-type: none"> • Documents officiels relatifs aux politiques, plans de mise en œuvre du programme et rapports • Statistiques officielles • Données de suivi du programme • Dossiers du programme 	<ul style="list-style-type: none"> • Examen des documents relatifs à la planification du programme, comptes rendus de réunions, rapports d'étape • Situation politique, socio-économique et/ou sanitaire du pays ou du lieu spécifique dans lequel le programme a été mis en œuvre
Collecte de données auprès de groupes ou d'individus	<ul style="list-style-type: none"> • Entretiens⁴ avec des informateurs clés, individus, groupes, groupes de discussion, méthodes projectives • Questionnaires ou enquêtes : par courriel, sur Internet, en face à face, données mobiles • Méthodes spécialisées (p. ex., dotmocracy [méthode participative d'aide à la décision], classement hiérarchique par cartes, calendriers saisonniers, méthodes projectives, histoires⁵) 	<ul style="list-style-type: none"> • Entretiens d'informateurs clés avec des représentants des ministères compétents, organisations non gouvernementales et/ou la communauté de développement au sens large • Entretiens avec des responsables, agents d'exécution et personnes chargées du suivi régulier du programme • Entretiens, discussions de groupe (telles que des groupes de discussion) et/ou questionnaires auprès de participants au programme

³ De plus amples informations sur chacun de ces éléments et une liste plus complète des options de collecte/regroupement de données sont disponibles sur la page « Collect and/or Retrieve Data » du site Internet de BetterEvaluation, à l'adresse suivante http://betterevaluation.org/plan/describe/collect_retrieve_data (en anglais).

⁴ Voir Note N°12, Entretiens.

⁵ La *dotmocracy* permet de recueillir les niveaux d'accord sur des déclarations écrites auprès d'un grand nombre de personnes. Le classement hiérarchique par cartes fournit un aperçu de la manière dont chacun classe et hiérarchise différents phénomènes. Les calendriers saisonniers permettent de visualiser des tendances de variations sur certaines périodes de temps. Les méthodes projectives fournissent une amorce pour les entretiens (p. ex., à l'aide du photolangage : les participants sélectionnent une ou deux images dans une série d'images et les utilisent pour illustrer leurs commentaires sur un sujet). Les histoires, personnelles par exemple, donnent un aperçu de la manière dont les gens vivent.

Observation	<ul style="list-style-type: none"> • Structurée ou non structurée • Avec ou sans participants • Participative ou non participative • Enregistrée à l'aide de notes, photos ou vidéos 	<ul style="list-style-type: none"> • Observations des activités du programme et des interactions avec les participants
Mesures physiques	<ul style="list-style-type: none"> • Mesures biophysiques • Informations géographiques 	<ul style="list-style-type: none"> • Poids des nourrissons • Sites fortement touchés par l'infection à VIH

Utiliser un éventail de méthodes de collecte et d'analyse de données

De nombreuses évaluations d'impact utilisent une variété de méthodes. Mais « l'évaluation à méthodes mixtes » se distingue par *l'intégration systématique* de méthodes et méthodologies quantitatives et qualitatives à tous les stades de l'évaluation⁶. L'emploi de méthodes mixtes s'explique notamment par le fait qu'elles permettent de surmonter les points faibles inhérents à chaque méthode utilisée lorsqu'appliquée de manière isolée. Des informations convergentes (cohérentes vis-à-vis de l'orientation des constatations) provenant de différentes sources de données renforcent en outre la crédibilité des résultats de l'évaluation et permettent de mieux comprendre le programme/la politique, ses effets et son contexte⁷.

Pour décider de l'utilisation d'une approche de méthodes mixtes, il convient de déterminer :

- à quelle étape de l'évaluation utiliser des méthodes mixtes (le modèle est considéré comme plus puissant si les méthodes mixtes sont intégrées à plusieurs étapes de l'évaluation) ;
- si les méthodes sont utilisées de manière consécutive (les données d'une source servent de base à la collecte de données auprès d'une autre) ou simultanée (on utilise la triangulation pour comparer les informations provenant de différentes sources indépendantes) ;
- si les méthodes qualitatives et quantitatives utilisées auront la même pondération ou non ;
- si le modèle concerne un seul niveau (p. ex., les ménages) ou plusieurs (p. ex., un programme national qui nécessite une description et une analyse des liens entre les différents niveaux).

Le cadre analytique spécifique et le choix des méthodes d'analyse des données particulières dépendent du but de l'évaluation d'impact et du type de questions clés d'évaluation qui y sont liées de manière intrinsèque :

- les **questions descriptives** requièrent des méthodes d'analyse de données prenant en compte des données quantitatives et qualitatives ;
- les **questions causales** nécessitent l'emploi d'un modèle de recherche qui tient compte de l'attribution (les changements observés sont-ils dus à l'intervention ou à des facteurs

⁶ Bamberger Michael, « Introduction aux méthodes mixtes dans l'évaluation d'impact », Note sur l'évaluation d'impact n° 3, InterAction, Washington, août 2012. Voir <http://www.interaction.org/impact-evaluation-notes>.

⁷ Ibid.

externes ?) et la contribution (dans quelle mesure l'intervention a-t-elle induit les changements observés ? ; voir Note N°6, Présentation des stratégies d'attribution causale) ;

- les **questions évaluatives** demandent des stratégies de synthèse qui appliquent les critères d'évaluation aux données afin de répondre aux questions clés d'évaluation (voir Note N°3, Critères d'évaluation). Le fait de définir dès le début ce qui constitue une « réussite » en mettant au point des grilles d'évaluation critériées spécifiques (normes ou niveaux de performances du programme ou de la politique) fournit une base à laquelle il est possible d'associer systématiquement les informations recueillies, afin de porter des jugements transparents et fondés sur des données probantes quant à la valeur du programme ou de la politique (également appelée « logique d'évaluation », voir Note N°4, Logique d'évaluation).

Bien qu'une évaluation d'impact vise à examiner les résultats à long terme d'un programme ou d'une politique, les décideurs ont souvent besoin d'informations plus rapides, et il est par conséquent nécessaire de recueillir également des informations sur les résultats à court terme. L'on sait par exemple que les résultats des interventions dans le domaine de l'éducation n'apparaissent que tardivement : l'évaluation de l'initiative des écoles amies des enfants à Moldova a ainsi capturé les résultats à court terme (comme « une participation accrue des élèves à l'apprentissage par le biais de méthodes d'enseignement interactives et participatives »⁸) mesurés lors de l'intervention ou peu après son achèvement et a présumé que ces résultats étaient annonciateurs d'effets à plus long terme.

Cependant, constater simplement un changement – en mesurant les indicateurs clés – ne permet pas de l'expliquer. Il faut également recueillir des informations sur des activités spécifiques réalisées, et sur le contexte dans lequel elles ont été mises en œuvre. Comme précisé plus haut, il est essentiel de disposer d'une théorie du changement explicite pour le programme ou la politique afin d'identifier les mesures qui doivent être collectées. Cette théorie fournit en outre une orientation en précisant les aspects de la mise en œuvre du programme – et de son contexte – sur lesquels la collecte de données doit être axée. En précisant en amont le cadre d'analyse des données, les besoins spécifiques pour la collecte de données (données primaires ou nouvelles) et le regroupement de données (données secondaires ou existantes) sont clairement intégrés d'une manière qui montre également comment les données seront analysées afin de répondre aux questions clés d'évaluation et d'aboutir à des jugements évaluatifs. Les besoins en matière de données et les méthodes de collecte et d'analyse des données correspondant à chacune des questions clé d'évaluation doivent être décrits dans le plan d'évaluation, tout comme les détails de leur collecte (comment, où, quand et auprès de qui), en fonction de la stratégie d'échantillonnage de la population étudiée, des sites et/ou des périodes de temps.

S'assurer que les méthodes de collecte et d'analyse de données sont réalisables

Une fois la planification terminée, il est important de vérifier la faisabilité des méthodes de collecte et d'analyse de données pour s'assurer que les éléments proposés peuvent effectivement être réalisés pendant la durée de l'évaluation et avec les ressources disponibles. Il se peut par exemple que les informateurs clés ne soient pas disponibles au moment où les évaluateurs auront besoin de données. Il est également capital d'étudier l'équipement et les compétences indispensables pour recourir à ces méthodes, et de déterminer s'ils sont disponibles ou s'ils peuvent être obtenus ou développés. Par exemple, en cas de collecte de données de questionnaire par téléphone portable, il est indispensable que chaque collecteur de données dispose de ce type de matériel ou qu'il existe un système fiable de partage de téléphones portables dans l'équipe. Il convient de pallier les importantes lacunes entre les

⁸ Velea, Simona, et CReDO (centre d'information sur les droits de l'homme), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011), Republic of Moldova*, ministère de l'Éducation de la République de Moldova/UNICEF, 2012. Voir http://www.unicef.org/moldova/CFS_EN_PRINT.pdf (en anglais).

équipements disponibles et les équipements nécessaires en se procurant des ressources supplémentaires ou, de façon plus réaliste, en adaptant les méthodes aux ressources disponibles.

Comme il est impossible de tout prévoir, et que certaines conditions peuvent évoluer au cours de l'évaluation, il peut s'avérer nécessaire de revoir les options choisies et de modifier le plan d'évaluation en conséquence. Dans ce cas, il est important de recenser ce qui a changé et pourquoi, d'examiner et d'indiquer toute incidence sur le produit de l'évaluation et son utilisation.

3. ASSURER UNE BONNE GESTION DES DONNÉES

Une bonne gestion des données suppose la mise en place de processus efficaces pour leur collecte et enregistrement cohérents, leur stockage sécurisé, leur [nettoyage](#), leur transfert (p. ex., entre les différents types de logiciels utilisés pour l'analyse), leur présentation efficace et leur accessibilité à des fins de vérification et d'utilisation par autrui.

Certains des aspects de la qualité des données fréquemment cités sont :

- la **validité** : les données mesurent ce qu'elles sont censées mesurer.
- la **fiabilité** : les données sont mesurées et collectées de manière cohérente en fonction de définitions et de méthodes normalisées ; la répétition des mesures produit les mêmes résultats.
- l'**exhaustivité** : tous les éléments de données sont inclus (conformément aux définitions et méthodes spécifiées).
- la **précision** : les données sont suffisamment détaillées⁹.
- l'**intégrité** : les données sont protégées de tout biais ou manipulation délibérée pour des raisons politiques ou personnelles.
- l'**actualité** : les données sont à jour (actuelles) et les informations sont disponibles à temps.

Il est conseillé d'utiliser des outils de collecte de données normalisés, qui ont déjà été testés et éprouvés dans des conditions réelles, et de les améliorer si nécessaire afin d'optimiser la qualité des données. Lorsqu'il faut les adapter au contexte local, ou lorsque des outils de collecte de données doivent être développés, il est important de procéder en premier lieu à un [essai pilote](#) (et d'améliorer l'outil) avant de l'utiliser de manière plus générale. Avoir recours à des collecteurs de données expérimentés, les former à une tâche ou à un outil spécifique et/ou superviser la collecte de données de plusieurs collecteurs peut également aider à réduire la partialité (p. ex. encourager de façon inappropriée certaines réponses pendant les entretiens) ou les erreurs (p. ex. une mauvaise compréhension des éléments du programme devant être étudiés) dans les données obtenues.

Les évaluateurs ne sont pas nécessairement les seuls responsables de la collecte de données. De nombreuses études mentionnent les avantages de l'« [évaluation participative](#) », qui peut aller au-delà de la simple collecte de données et impliquer des membres du personnel du programme, des participants et/ou d'autres parties prenantes pour : fixer le calendrier de l'évaluation, identifier les résultats clés et définir ce qui constitue une « réussite », contribuer à collecter des données, et analyser et interpréter les résultats (voir Note N°5, Approches participatives).

Même lorsque les données ont été recueillies à l'aide de procédures clairement définies et d'outils normalisés, il est nécessaire de vérifier qu'aucune donnée n'est inexacte ou manquante. Ce

⁹ Définition plus précise de mesures quantitatives : une mesure est considérée comme valide si elle est à la fois exacte et précise. L'exactitude est définie comme l'écart par rapport à la valeur « réelle » et la précision comme la « dispersion ». Autrement dit, l'exactitude indique à quel point la mesure prise est proche de la valeur effective (réelle) ; la précision correspond au degré auquel des mesures répétées dans des conditions inchangées aboutissent aux mêmes résultats.

processus s'appelle le nettoyage des données, et il consiste à détecter et à traiter toutes les erreurs qui se produisent lors de l'écriture, de la lecture, du stockage, de la transmission ou du traitement des données informatisées.

Assurer la qualité des données signifie également garantir des analyses et une présentation des données appropriées dans le rapport d'évaluation avec par conséquent des résultats clairs et des conclusions étayées. Il est par ailleurs nécessaire de rendre les données accessibles afin qu'elles puissent être vérifiées par d'autres et/ou utilisées à d'autres fins, par exemple pour synthétiser les résultats de différentes évaluations ([examen systématique](#), [méta-analyse](#), [examen réaliste](#) ou autre [méta-évaluation](#)).

4. QUESTIONS SPÉCIFIQUES POUR GARANTIR LA QUALITÉ DE LA COLLECTE DE DONNÉES

Comme indiqué ci-dessus, garantir la qualité de la collecte des données fait *partie* de la bonne gestion des données. Quelques questions particulièrement pertinentes sont abordées plus en détail dans cette section.

Rendre la stratégie d'échantillonnage explicite

La stratégie d'échantillonnage utilisée est un élément particulièrement important qui doit être décrit en détail dans le plan et le rapport de l'évaluation d'impact. En premier lieu, il convient de définir la « population à étudier » pour l'échantillon. Il s'agit du groupe ou des « unités » à étudier (tous les couples mariés par exemple) dans la zone géographique visée (comme un village particulier) au cours de la période cible (par exemple les mariages enregistrés depuis 1995). L'échantillonnage est le processus de sélection des unités (p. ex. les individus, les groupes) de la population à étudier afin d'analyser ces unités en détail dans le but de tirer des conclusions sur la population générale (il est à noter que la généralisation des constatations de l'évaluation dépendra de la façon dont l'échantillonnage est réalisé). [L'erreur d'échantillonnage](#) est propre à n'importe quel échantillon. En d'autres termes, aucun échantillon ne permet d'obtenir exactement les mêmes informations que si toutes les personnes dans la population à étudier ont été incluses dans la collecte de données. Par conséquent, les informations recueillies auprès d'un échantillon servent à réaliser des *estimations* concernant la population étudiée. Le tableau 3 fournit un aperçu des trois groupes de base d'options d'échantillonnage.

Tableau 3. Aperçu des options d'échantillonnage avec méthodes illustratives

Groupe d'options d'échantillonnage	Méthodes spécifiques	Risque d'introduire un biais
Probabilité : utilisation de méthodes aléatoires ou quasi aléatoires pour sélectionner l'échantillon, puis emploi de la généralisation statistique afin de faire des déductions sur cette population	<ul style="list-style-type: none"> • Aléatoire simple • Aléatoire stratifié • Multi-étapes • Séquentiel 	Pour ce groupe, il existe des règles spécifiques concernant la sélection du cadre d'échantillonnage, la taille de l'échantillon et la gestion des variations au sein de l'échantillon.

<p>Choix raisonné : étude de cas riches en informations sur une population donnée afin de faire des déductions analytiques concernant la population. Les unités sont sélectionnées en fonction d'une ou de plusieurs caractéristiques prédéterminées et l'échantillon peut être constitué d'une seule unité.</p>	<ul style="list-style-type: none"> • Confirmation et infirmation • Cas critique • Variation maximale • Valeur aberrante • Boule de neige • Fondé sur la théorie • Cas typique 	<p>Ce groupe encourage la transparence dans la sélection et la triangulation des cas, et l'infirmation d'éléments de preuve</p>
<p>Convenance : ces options d'échantillonnage s'appuient sur des individus qui sont disponibles ou des cas dès qu'ils se produisent.</p>	<ul style="list-style-type: none"> • Facilement accessibles • Bénévoles 	<p>Ce type d'échantillonnage a la plus faible crédibilité, mais nécessite moins d'investissements en termes de temps, d'argent et d'efforts</p>

Lors de l'évaluation des résultats d'un programme ou d'une politique, différentes méthodes d'échantillonnage introduiront différents types de biais. Ainsi, il est important de décrire clairement les points forts et les points faibles de la stratégie d'échantillonnage utilisée et de les aborder, dans la mesure du possible, dans l'analyse des données et/ou de décrire leur incidence sur les conclusions tirées au sujet du programme/de la politique évalué(e).

Trois groupes d'options d'échantillonnage sont décrits ci-dessus ; il est toutefois conseillé d'éviter d'utiliser l'échantillonnage de « convenance » car il ne permet pas de généraliser correctement les résultats (c'est-à-dire, faire des estimations concernant la population générale à partir des résultats relatifs à l'échantillon étudié). Dans un échantillonnage aléatoire, chaque unité a la même probabilité ou une probabilité connue, ce qui augmente l'éventualité que les informations recueillies à propos de l'échantillon soient représentatives de l'ensemble de la population à étudier, en particulier si les conditions de taille de l'échantillon sont remplies (voir Note N°7, Essais contrôlés randomisés). Ce type d'échantillonnage permet une « généralisation statistique » dans laquelle une certaine marge d'erreur est définie (en général entre 5 et 10 %) qui détermine le niveau de risque acceptable d'erreur quant aux résultats du programme. L'échantillonnage raisonné n'est pas basé sur la représentation ou le caractère aléatoire. Il est utilisé pour sélectionner les unités qui permettent d'en apprendre beaucoup sur les questions qui sont importantes pour l'évaluation (tels que les « informateurs clés »)¹⁰. Ce type d'échantillonnage permet une « généralisation analytique » (projections de la probable transférabilité des constatations basées sur une analyse théorique de l'effet du contexte et des facteurs qui produisent les effets directs).

Gérer les abandons et les refus

Un échantillonnage approprié peut aider à réduire le « biais de sélection ». Toutefois, des taux élevés d'abandon (ou « perdus de vue » nombreux) peuvent être à l'origine d'un certain biais : un grand nombre de personnes quittent le programme, et il n'est plus possible de collecter des données auprès d'elles. De même, un faible taux de réponse à un questionnaire ou un nombre élevé de personnes refusant de participer à l'effort de collecte de données peuvent introduire des « biais de mesure ». Il

¹⁰ Taylor-Powell, Ellen, *Sampling*, University of Wisconsin, Milwaukee, 1998.

existe plusieurs moyens pour résoudre ces problèmes, notamment : déployer des efforts spécifiques pour accroître la participation ; collecter des informations auprès de plusieurs sources, informations qui peuvent ensuite être triangulées ; déterminer en quoi les personnes qui participent sont différentes de celles qui abandonnent ou refusent tout simplement de participer et aussi détecter les similitudes entre les participants et la population à étudier (et, ensuite, évaluer l'incidence sur l'interprétation et la généralisation des résultats). Encore une fois, la transparence et des discussions approfondies sont nécessaires.

5. QUESTIONS SPÉCIFIQUES POUR GARANTIR LA QUALITÉ DE L'ANALYSE DE DONNÉES

Comme pour la collecte de données, garantir la qualité de l'analyse des données fait *partie* de la bonne gestion des données. Quelques questions particulièrement importantes sont abordées plus en détail dans cette section.

Chaque source de données a ses points forts et ses faiblesses, qui doivent être décrits de manière explicite (en termes de qualité des données, voir ci-dessous) et clairement compris afin de s'assurer que les sources sont utilisées et interprétées de façon appropriée (voir la section 7 ci-dessous). Ainsi, l'intégration délibérée de différentes méthodes de collecte de données et/ou la triangulation de différentes sources de données pour répondre aux questions clés d'évaluation permet de surmonter les points faibles inhérents à chaque méthode lorsqu'utilisée seule. Des méthodes mixtes renforcent en outre la crédibilité des résultats si des informations provenant de différentes sources de données se rejoignent (cohérentes vis-à-vis de l'orientation des résultats). Elles permettent en outre de mieux comprendre un programme ou une politique, ses effets et son contexte. Même s'il est possible d'associer les méthodes ultérieurement (après la collecte des données), il est conseillé, et souvent beaucoup plus facile, de le faire en amont (pendant la planification de l'évaluation). Ainsi, la collecte des données peut spécifiquement viser à garantir que la bonne combinaison de données probantes est disponible au bon moment pour répondre aux questions clés d'évaluation. Le tableau 4 met en évidence les principaux objectifs de la combinaison de sources de données.

Tableau 4. L'intérêt d'associer des méthodes : principaux objectifs de la combinaison de sources de données

Enrichir	Utiliser des données qualitatives pour identifier les problèmes ou recueillir des informations sur les variables qui ne peuvent pas être obtenues avec des approches quantitatives
Étudier	Générer des hypothèses à partir de données qualitatives afin qu'elles soient vérifiées à l'aide des données quantitatives (telles que l'identification des sous-groupes devant être analysés séparément dans les données quantitatives, p. ex. pour étudier l'impact différencié)
Expliquer	Utiliser des données qualitatives pour comprendre les résultats imprévus des données quantitatives
Triangler (confirmer ou rejeter)	Vérifier ou rejeter les résultats de données quantitatives à l'aide de données qualitatives (ou inversement)

Analyser des données pour les synthétiser et rechercher des tendances est une étape importante de toute évaluation. Les stratégies d'analyse des données et méthodes de synthèse de données doivent être définies lors de la conception de l'évaluation, être adaptées au type de questions clés d'évaluation et entièrement décrites dans le plan d'évaluation.

Diverses méthodes d'analyse sont disponibles pour les questions clés d'évaluation (voir tableau 5). Ces options peuvent être regroupées en deux catégories principales : les options pour les données quantitatives (chiffres) et les options pour les données qualitatives (p. ex. du texte).

Tableau 5. Méthodes pour l'analyse des données numériques et textuelles

Analyse numérique

Analyse de données numériques, telles que le coût, la fréquence ou des caractéristiques physiques. Les méthodes possibles sont notamment

- Corrélation : une technique statistique pour déterminer l'importance des liens associant deux variables ou plus.
- Croisements : permettent d'obtenir une indication sur la concomitance de deux variables (p. ex. le sexe et la fréquence de scolarisation).
- Exploration de données et de texte : techniques automatisées et informatisées qui exploitent de grandes quantités de texte ou de données afin de trouver de nouvelles tendances et informations.
- Techniques exploratoires : premier examen rapide de l'ensemble des données qui en résume les principales caractéristiques, souvent grâce à l'utilisation des méthodes visuelles.
- Tableaux de fréquence : organisation des données collectées par ordre croissant en fonction de leur importance, avec leur fréquence d'apparition, afin de garantir une image plus claire de l'ensemble de données.
- Mesure de la médiane : mesure synthétique qui vise à décrire un ensemble de données avec une valeur unique qui représente le milieu ou centre de sa distribution.
- Indice de dispersion : mesure synthétique qui décrit la distribution des valeurs autour du centre.
- Analyse descriptive multivariée : elle fournit des résumés simples de (grandes quantités d') informations (ou de données) avec deux ou plusieurs variables connexes.
- Inférence non paramétrique : les données sont flexibles et ne suivent pas une distribution normale.
- Inférence paramétrique : effectuée sur des données qui répondent à certains paramètres. Les données seront normales (la distribution est parallèle à la courbe en cloche) ; les chiffres peuvent être additionnés, soustraits, multipliés ou divisés ; les variances sont égales lors de la comparaison de deux groupes ou plus ; et l'échantillon doit être important et sélectionné de façon aléatoire.
- Statistiques de synthèse : elles fournissent un résumé rapide des données, particulièrement utile pour comparer un projet à un autre, l'avant et l'après.
- Analyse de séries chronologiques : observation d'éléments de données bien définis obtenus à l'aide de mesures répétées dans le temps.

Analyse textuelle

Analyse de mots, parlés ou écrits, y compris les réponses au questionnaire, entretiens et documents. Les méthodes possibles sont notamment :

- Analyse de contenu : condenser des quantités importantes de contenu textuel non structuré en données gérables correspondant aux questions de recherche (d'évaluation).
- Codage thématique : enregistrement ou identification de passages de texte ou d'images en rapport avec un thème commun ou idée commune, permettant le classement du texte en catégories.
- Textes explicatifs : rédaction de récits cohérents des changements vécus par un individu, une communauté, un site, un programme ou une politique.
- Calendrier : liste des événements clés classés par ordre chronologique.

Concernant les questions clés d'évaluation causales, il existe essentiellement trois grandes approches de l'analyse d'attribution causale : (1) les approches contrefactuelles (2) la cohérence entre données probantes et relation de cause à effet et (3) l'exclusion d'autres explications (voir tableau 6). Dans l'idéal, il convient de combiner ces trois approches pour établir une causalité (voir Note N°6, Présentation des stratégies d'attribution causale).

Pour les questions évaluatives, il convient d'employer des grilles d'évaluation critériées spécifiques afférentes aux critères d'évaluation utilisés (critères de l'OCDE-CAD par exemple) pour synthétiser les données probantes et apprécier la valeur du programme ou de la politique (voir Note N°4, Logique d'évaluation).

Tableau 6. Approches de l'analyse d'attribution causale

Approches contrefactuelles	Estimation de ce qui se serait passé en l'absence du programme ou de la politique, ce qui implique d'avoir recours à un groupe contrôle ou groupe témoin.
Cohérence entre données probantes et relation de cause à effet	Identification des tendances cohérente avec une relation de cause à effet, puis confirmation ou infirmation des données probantes.
Exclusion d'autres explications	Identification de possibles autres explications causales, puis recherche d'informations permettant de déterminer si ces explications peuvent être exclues.

La visualisation de données est un outil puissant pour analyser et communiquer les résultats de l'évaluation. Ce processus consiste à représenter des données sous forme graphique afin d'identifier des tendances et modèles qui, autrement, seraient confus ou difficiles à discerner. Il existe de nombreuses possibilités¹¹, et le choix du type de visuel à utiliser dépend de la nature des données (p. ex. relationnelles, comparatives, temporelles). En outre, le type de visuel utilisé pour faciliter

¹¹ Pour de plus amples informations sur les méthodes de visualisation de données, voir la page Internet « Visualise Data », BetterEvaluation, http://betterevaluation.org/plan/describe/visualize_data (en anglais).

l'analyse de données (p. ex. les pics de trafic d'un site Internet associé à une campagne sur des médias sociaux) n'est pas nécessairement le plus efficace à des fins de communication. Lors de la diffusion des données, il peut être indispensable de simplifier le visuel et de mettre en exergue les zones déterminantes afin d'attirer l'attention des lecteurs et des parties prenantes. Les conclusions de l'évaluation d'impact doivent s'appuyer sur les données et être présentées de manière transparente et vérifiable par les lecteurs.

6. QUESTIONS ÉTHIQUES ET LIMITES PRATIQUES

Questions éthiques

Toutes les évaluations doivent respecter les *Ethical Guidelines for Evaluation*¹² (directives éthiques pour l'évaluation) définies par le Groupe des Nations Unies pour l'évaluation (GNUE) et auxquelles adhère l'UNICEF. Ces directives fixent un code de conduite pour les évaluateurs ainsi que des obligations d'ordre éthique relatives à la collecte et au stockage de données pour les personnes participant à l'évaluation (voir également la Note N°1, Présentation de l'évaluation d'impact). Ces obligations concernent notamment : le respect de la dignité et de la diversité, les droits, la confidentialité et la prévention des préjudices. Elles doivent être abordées de façon explicite dans le modèle d'évaluation et dans l'ensemble du processus d'évaluation (consulter le [site Internet Ethical Research Involving Children](#))¹³.

Dans la plupart des cas, un comité institutionnel d'éthique doit officiellement approuver le mode de collecte des données ainsi que leur gestion et utilisation dans le cadre de l'évaluation. Il s'agit de veiller à ce que ces pratiques soient conformes aux obligations éthiques de toutes les personnes auprès desquelles des données sont recueillies.

L'évaluation de l'initiative moldave des écoles amies des enfants¹⁴ a fourni quelques bons exemples de stratégies pour aborder les questions éthiques lors de la collecte et de l'analyse de données.

- Au cours de la phase préparatoire, les parties prenantes (UNICEF et partenaires de mise en œuvre) ont été consultées, puis les outils de collecte de données ont été testés sur un petit groupe d'enseignants, d'enfants et de parents qui ont accepté de participer après avoir été informés sur le processus. Cette démarche avait pour objectif d'identifier tous les éléments pouvant être considérés comme provoquant des nuisances, des situations inconfortables ou affectant les droits des participants.
- Au cours de la phase de collecte des données, il a été précisé aux répondants que toutes les données demeureraient confidentielles et seraient utilisées à des fins d'évaluation exclusivement. Ce point était également précisé dans le texte d'introduction des questionnaires, qui ne comportaient par ailleurs aucun champ consacré aux données personnelles pouvant faciliter l'identification des répondants.
- Lors des réunions des groupes de discussion, les noms des participants n'étaient pas consignés et les sessions n'étaient pas enregistrées. La participation reposait sur un

¹² Groupe des Nations Unies pour l'évaluation, *Ethical Guidelines for Evaluation*, GNUE, New York, 2007.

¹³ Le site Internet www.childethics.com est un référentiel d'informations basées sur des données probantes, des ressources et des liens vers des articles de journaux collectés par le projet Ethical Research Involving Children afin d'orienter et d'améliorer la recherche impliquant les enfants et de fournir une plateforme pour une plus ample réflexion et un dialogue critiques.

¹⁴ Velea, Simona et CREDO (centre d'information sur les droits de l'homme), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011)*, Republic of Moldova, ministère de l'Éducation de la République de Moldova/UNICEF, 2012.

consentement éclairé : les participants avaient reçu des informations sur l'approche de l'évaluation, leur rôle dans cette évaluation et les avantages qu'ils pouvaient en tirer à titre personnel, directement (expérience d'apprentissage par exemple) et indirectement (en contribuant à une meilleure compréhension du système éducatif).

- Pendant les entretiens, les échanges entre participants et ceux entre les participants et l'évaluateur reposaient sur le respect mutuel et la confiance.

Cette évaluation moldave apporte également des exemples intéressants de mesures de protection destinées à assurer la confidentialité des données :

- Lors du traitement des données, il n'a pas été fait référence à des écoles en particulier ou à des déclarations faites spécifiquement par certains élèves, enseignants ou parents.
- Les données provenant des questionnaires ont fait l'objet d'un traitement statistique et les données qualitatives ont été étudiées dans leur ensemble, en l'absence de toute information personnelle. (Les commentaires spécifiques à propos d'écoles particulières identifiées ont toutefois été repris.)

Limites pratiques

Les limites pratiques fréquemment rencontrées dans la collecte de données sont les suivantes :

- les [informateurs clés](#), participants et sites concernés sur le terrain peuvent ne pas être disponibles/être injoignables si
 - les informations concernant les sites et/ou les participants du programme sont incomplètes,
 - ils ne sont pas contactés suffisamment tôt avant ou pendant la période de collecte de données,
 - l'évaluation est réalisée après la fin du programme ;
- les sources de données pertinentes existantes (telles que les indicateurs pour les rapports intermédiaires élaborés par les partenaires d'exécution, les données d'enquête) sont limitées en termes de dimension temporelle (p. ex. aucune référence, aucune série chronologique disponible) et/ou de contenu (il manque par exemple des indicateurs importants) et/ou de qualité (il existe des problèmes importants quant à la qualité des données) ;
- les informations relatives à la qualité des services fournis par différents agents d'exécution dans le cas d'un programme à grande échelle sont généralement limitées.

Si les ressources et le calendrier de l'évaluation le permettent, certaines de ces limites concernant la collecte des données peuvent être corrigées (p. ex. en prolongeant la période ou en étendant le champ ou la portée de la collecte de données). Mais il faudra probablement tirer des conclusions à partir de données imparfaites. Par conséquent, il est important de préciser ouvertement les limites et de décrire comment elles ont pu affecter les résultats, les conclusions et les recommandations de l'évaluation.

7. EXEMPLE DE BONNES PRATIQUES

L'évaluation des écoles amies des enfants¹⁵ susmentionnée a fourni quelques bons exemples de réflexion bien structurée sur les méthodes de collecte et d'analyse de données et sur leur mise en œuvre (voir encadré 2).

Identification des besoins en matière de données par une décomposition appropriée en concepts clés :

La focalisation sur l'enfant était définie comme une approche méthodologique pouvant être « décomposée » à partir des éléments constitutifs suivants :

- connaître les styles d'apprentissage, problèmes et aspirations des élèves ;
- faire participer et valoriser tous les élèves, indépendamment de leur origine ethnique, statut social, handicap, religion ou sexe ;
- ne plus se centrer sur l'enseignement pour se concentrer sur l'apprentissage et orienter l'ensemble du processus d'enseignement sur l'enfant ;
- faire participer les élèves à la mise en place d'objectifs, à l'apprentissage de nouvelles méthodes d'étude et à l'évaluation des performances (mise en place et communication des critères) et à l'auto-évaluation ;
- s'assurer que l'on enseigne aux élèves comment apprendre et que ce qu'ils apprennent présente un intérêt dans leur vie quotidienne, en utilisant les expériences d'apprentissage positives acquises par les élèves à l'extérieur de l'école ;
- diversifier les cas de situations d'apprentissage et le rôle des enseignants : transmettre les informations, faciliter leur propre recherche d'informations, fournir des conseils, faire preuve de modération ;
- entretenir une bonne communication avec les parents et les encourager à participer à l'éducation de leur enfant.

La décomposition de ce concept central a ensuite amené les évaluateurs à constater le besoin de divers types de données, et par conséquent de différentes sources de données. Les données pertinentes ont été recueillies notamment à l'aide de questionnaires. La complexité de la focalisation sur l'enfant ne permettait pas d'utiliser seulement quelques indicateurs simples pour évaluer les résultats obtenus. Il a donc été nécessaire de grouper plusieurs mesures (obtenues par le biais d'une série de questions spécifiques dans l'instrument de collecte de données ; voir notes de bas de page dans le tableau 7 ci-dessous, tiré du rapport d'évaluation) liées aux domaines de résultats clés (tels que les conditions scolaires propices à l'égalité de traitement, voir les points énumérés dans le tableau 7).

¹⁵ Velea, Simona, et CReDO (centre d'information sur les droits de l'homme), *Child-Friendly Schools, External Evaluation Report of the Child-Friendly School Initiative (2007–2011)*, Republic of Moldova, ministère de l'Éducation de la République de Moldova/UNICEF, 2012, p. 47 et 48.

Tableau 7. Avis des élèves sur divers indicateurs relatifs à l'enseignement (moyennes ; comparaison de colonnes, score inférieur = meilleure position de l'indicateur donné.) Tiré du rapport d'évaluation sur les écoles amies des enfants en Moldavie, tableau 8, p. 47.

Items	Experimental schools	Control schools
1. Encouraging thinking in students, and self-assessment of their own learning styles ³¹	3.77	4.33
2. Setting hours and homework assignments together with the students ³²	15.48	16.89
3. Working with children who have disabilities, and those with learning difficulties ³³	7.74	8.51
4. Working with children of ethnic minorities ³⁴	6.7	7.33
5. Treating other teachers and all students with equal respect ³⁵	2.79	3.05
6. Creating school conditions favourable to equal treatment ³⁶	6.62	7.3
7. Encouraging communication and dialogue with other teachers and with parents and students ³⁷	25.02	26.24
8. Management of knowledge (correct learning, use of information technologies, etc.). ³⁸	5.1	5.31

31 Student questionnaire, Item 1, questions 10.4, 11.4, 11.12.

32 Ibid., Item 2, questions 9.1, 9.4, 9.3, 9.7, 9.8, 11.1, 11.7, 12.3, 12.4.

33 Ibid., Item 3, questions 11.3, 10.6, 11.12, 11.17.

34 Ibid., Item 4, questions 10.4, 11.4, 11.12.

35 Ibid., Item 5, questions 10.7, 11.7.

36 Ibid., Item 6, questions: 11.9, 11.11, 11.12, 11.17.

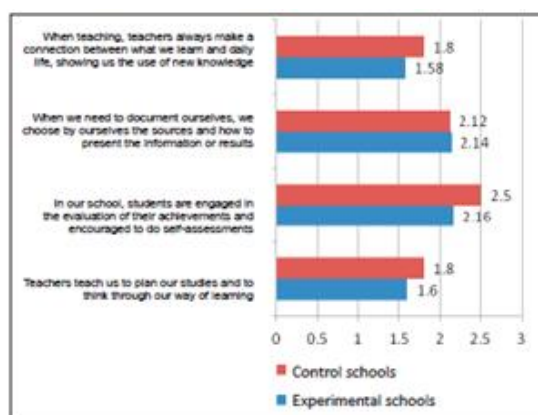
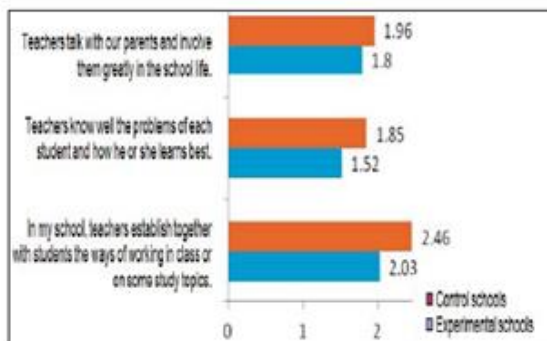
37 Ibid., Item 7, questions: 9.1-9.9, 11.1, 11.2, 11.11.

38 Ibid., Item 8, questions: 12.5, 12.1, 11.8.

Utilisation de différentes méthodes de collecte de données dans le but de trianguler diverses perspectives, et présentation claire des résultats :

L'évaluation a non seulement recueilli des données au sujet de la focalisation sur l'enfant auprès des enseignants, mais également sur les pratiques d'enseignement auprès de ceux qui en bénéficient : les élèves. Même si cette source de données peut sembler « évidente » pour certains évaluateurs, « s'adresser aux participants du programme » n'est pas une démarche systématique de toutes les évaluations d'impact ; lorsque c'est le cas, elle est souvent considérée comme purement « symbolique » (voir également la Note N°5, Approches participatives). Les données complexes présentées dans cette étude sur les écoles amies des enfants étaient très accessibles grâce à une bonne visualisation des données (voir les figures 15 et 16 ci-dessous, également tirées du rapport d'évaluation).

Figure 1. Avis des élèves concernant les pratiques d'enseignement pertinentes pour la focalisation sur l'enfant (figure de gauche) et avis des élèves concernant les pratiques d'enseignement (figure de droite). (Moyennes ; 0 = note maximale, 3 = note minimale). Tiré du rapport d'évaluation sur les écoles amies des enfants en Moldavie, Figure 15 (gauche) et Figure 16 (droite), p. 48.



Évaluer les impacts différentiels du programme

Outre les « moyennes » communiquées, les évaluateurs ont également étudié les impacts différentiels du programme. Ces derniers sont essentiels pour garantir une meilleure compréhension de ce qui a fonctionné pour qui et dans quelles conditions.

Les évaluateurs ont par exemple signalé des résultats positifs dans le comportement des élèves, illustrés par une participation plus importante et plus fréquente aux activités scolaires, en particulier dans le cas des élèves handicapés ou ceux appartenant à des groupes vulnérables (tels que définis dans l'étude). En outre, plus d'enfants ont obtenu de meilleurs résultats scolaires, avec des progrès notables réalisés par les élèves en difficulté (tels que définis dans l'étude).

8. EXEMPLES DE DIFFICULTÉS RECONTRÉES

Les *mauvais choix de méthodes* ou une *mise en œuvre insatisfaisante des méthodes* illustrent certaines des mauvaises pratiques dans la collecte et l'analyse de données. Quelques exemples sont proposés ci-dessous.

Voici quelques exemples de mauvais choix de méthodes :

1. Choisir toujours les mêmes méthodes, qu'elles soient adaptées ou non à l'évaluation concernée.

Comme le montre clairement cette note, de multiples méthodes de collecte et d'analyse des données peuvent être utilisées dans les évaluations d'impact. La combinaison la plus appropriée de méthodes dépend de la nature de l'objet de l'évaluation et de son contexte. Autrement dit, il convient de répondre aux questions suivantes : quelles sont les questions d'évaluation auxquelles il faut répondre ? À quel moment les différentes parties prenantes ont-elles besoin d'une réponse à ces questions pour étayer leurs décisions ? Quelles sont les ressources disponibles pour l'évaluation ? Quelles autres conditions peuvent entraver l'évaluation (p. ex. l'environnement politique) ?

2. Choisir des méthodes qui ne sont pas spécifiquement liées aux questions clés de l'évaluation.

Élaborer une matrice d'évaluation (voir encadré 1 plus haut) peut permettre de s'assurer que toutes les questions clés d'évaluation sont prises en compte.

3. Choisir une seule méthode pour répondre à une question clé d'évaluation.

Il ne s'agit pas simplement d'avoir recours à plusieurs méthodes, mais de les choisir de manière à ce que celles-ci se complètent, réduisant les risques de biais et garantissant la triangulation. En d'autres termes, les points forts d'une méthode sélectionnée pallient (au moins en partie) les faiblesses d'une autre, et les différentes perspectives sur l'intervention (en termes des parties prenantes clés) sont prises en compte.

Mise en œuvre insatisfaisante des méthodes

1. Mauvaise réalisation de l'échantillonnage.

Par exemple utiliser un échantillon de convenance ou avoir toujours un faible taux de réponse sans pour autant vérifier systématiquement et spécifier clairement en quoi les répondants diffèrent éventuellement des non-répondants. Concernant l'échantillonnage, le point le plus important est de veiller à ce que les données de l'échantillon puissent être correctement généralisées – statistiquement ou analytiquement. Une importante sous-représentation dans l'échantillon sélectionné peut conduire à des conclusions erronées ou incorrectes concernant le programme ou la politique évalué – par exemple sous-représentation de certains groupes d'âge, sexes ou autres données démographiques pertinentes, ou de certaines caractéristiques jugées en rapport avec les effets directs ou l'impact étudié (telles que l'exclusion des personnes ayant les résultats les moins bons dans le cas d'une évaluation d'impact des taux de réussite à des tests).

2. Collecte de données de mauvaise qualité.

Par exemple des questionnaires mal élaborés, l'incapacité à laisser aux participants l'occasion de réellement s'exprimer lors d'entretiens, l'absence de bons traducteurs ou interprètes et l'enregistrement incorrect des données. Il existe plusieurs solutions pour éviter ou minimiser les erreurs les plus fréquentes, dont notamment : obtenir l'avis de spécialistes concernant la conception ou l'adaptation des instruments de collecte de données ; garantir une collecte de données de bonne qualité par la sélection, la formation et la supervision des collecteurs de données (y compris en assurant l'équilibre entre les sexes au sein des équipes qui réalisent les entretiens) ; fournir des espaces privés pour les entretiens afin que les gens se sentent plus à l'aise pour exprimer leurs points de vue, ou utiliser des technologies permettant un meilleur « anonymat » dans les réponses et éviter d'obtenir des réponses « socialement souhaitables » ; et avoir recours à des traducteurs et interprètes agréés et expérimentés.

3. Analyse de données de mauvaise qualité.

Par exemple tirer des conclusions qui ne sont pas étayées par les données, ou se concentrer uniquement sur l'effet moyen du programme ou de la politique. Il existe plusieurs solutions pour permettre une analyse de données de bonne qualité : s'assurer que les conclusions « initiales » sont remises en question en organisant des sessions pour obtenir les commentaires des parties prenantes clés ou en cherchant volontairement à infirmer les données probantes ; rechercher des tendances ; et s'assurer que des questions « transversales » telles que l'équité sont traitées.

4. Informations et présentation de rapports de mauvaise qualité.

Par exemple en orientant les informations transmises en fonction des différents instruments de collecte de données (en reprenant les éléments du questionnaire un à un) au lieu de répondre aux questions clés d'évaluation, ou en utilisant des graphiques complexes qui doivent être amplement expliqués pour être compris. Une bonne visualisation des données, telle qu'une carte « visuelle » des principales idées concernant les données textuelles et l'ajout de citations particulières dans des encadrés afin d'illustrer les résultats sommaires, ou encore la mise en avant des principaux résultats à l'aide de feuilles de calcul soigneusement sélectionnées revêt une certaine importance pour communiquer les messages clés (de nombreuses options sont disponibles au-delà des paramètres par défaut).

9. LECTURES CLÉS ET LIENS UTILES

Bamberger, Michael, « Introduction aux méthodes mixtes dans l'évaluation d'impact », Note sur l'évaluation d'impact N°3, InterAction, Washington, août 2012. Voir <http://www.interaction.org/impact-evaluation-notes>.

BetterEvaluation, page Internet « Analyze Data », http://betterevaluation.org/plan/describe/look_for_patterns (en anglais).

BetterEvaluation, page Internet « Collect and/or Retrieve Data », http://betterevaluation.org/plan/describe/collect_retrieve_data (en anglais).

BetterEvaluation, page Internet « Combine Qualitative and Quantitative Data », http://betterevaluation.org/plan/describe/combining_qualitative_and_quantitative_data (en anglais).

BetterEvaluation, page Internet « Manage Data », http://betterevaluation.org/plan/describe/manage_data (en anglais).

Evergreen, Stephanie, D.H., *Presenting Data Effectively, Communicating Your Findings for Maximum Impact*, Sage, Thousand Oaks, 2013.

Measure Evaluation, page Internet « Data Quality Assurance », <http://www.cpc.unc.edu/measure/tools/monitoring-evaluation-systems/data-quality-assurance-tools/data-quality-assurance-tools> (en anglais).

Patton, Michael Quinn, *Qualitative Research & Evaluation Methods, troisième édition*, Sage, Thousand Oaks, 2001.

Perrin, Burt, « Faire le lien entre le suivi-évaluation et l'évaluation d'impact », Note sur l'évaluation d'impact N°2, InterAction, Washington, avril 2012. Voir <http://www.interaction.org/document/faire-le-lien-entre-le-suive-%C3%A9valuation-et-l-%C3%A9valuation-dimpact>.

GLOSSAIRE

<u>À plusieurs degrés</u>	<i>Échantillon aléatoire comportant plusieurs étapes (et souvent une étape d'échantillonnage en sous-groupes), telles que la sélection aléatoire de sous-groupes dans une population, puis la sélection aléatoire d'individus dans chacun des sous-groupes</i>
<u>Activité</u>	<i>Mesures prises ou processus à travers lesquels les intrants d'un projet sont mobilisés pour produire des résultats spécifiques. Par exemple des séances de conseil qui respectent les normes de qualité.</i>
<u>Données de référence</u>	<i>Données initialement collectées servant de base de comparaison avec les données recueillies ultérieurement. Par exemple les données recueillies avant une intervention à des fins de comparaison avec les données recueillies après la mise en œuvre de l'intervention.</i>
<u>Échantillonnage à variation maximale</u>	<i>Utilisé dans la recherche qualitative pour la sélection à dessein d'un grand nombre de cas.</i>
<u>Erreur d'échantillonnage</u>	<i>Mesure dans laquelle les résultats de l'échantillon diffèrent de ceux qui auraient été obtenus à partir de l'ensemble de la population, en raison d'une erreur aléatoire dans la sélection des répondants et de la perte de fiabilité associée.</i>
<u>Essai</u>	<i>Étude à petite échelle menée pour tester la fiabilité d'un outil de collecte de données.</i>
<u>Évaluation participative</u>	<i>Approche de l'évaluation qui implique les parties prenantes, en particulier les participants, d'un programme ou les personnes touchées par une politique donnée, dans certains aspects spécifiques du processus d'évaluation, dont la conception, la mise en œuvre de la collecte et de l'analyse des données et la supervision du travail des experts techniques.</i>
<u>Évaluation réaliste</u>	<i>Synthèse explicative et itérative de la recherche qui teste et construit des théories à moyenne portée. Elle fournit une analyse exploratoire et répond aux points suivants : ce qui fonctionne, pour qui, dans quelles circonstances, pourquoi et comment.</i>
<u>Extrant/Produit/Livrable</u>	<i>Effets immédiats, produits directs ou livrables des activités d'un programme/d'une politique. Par exemple le nombre de vaccins administrés</i>
<u>Impact</u>	<i>Effets à long terme, positifs et négatifs, primaires et secondaires, induits par une intervention, directement ou non, intentionnellement ou non¹⁶.</i>

¹⁶ Comité d'aide au développement de l'Organisation de coopération et de développement économiques, *Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats*, OCDE-CAD, Paris, 2010. Voir <http://www.oecd.org/development/peer-reviews/2754804.pdf>.

Note méthodologique n°10:
Présentation des méthodes de collecte et d'analyse de données dans l'évaluation d'impact

<u>Indicateur</u>	Mesure vérifiable choisie par l'équipe responsable d'un programme ou d'une politique afin de prendre des décisions concernant le programme/la politique. Par exemple la proportion d'élèves ayant réussi un examen normalisé.
<u>Informateur clé</u>	Participants à une recherche/évaluation susceptibles d'être bien informés sur un sujet et prêts à répondre sans parti pris.
<u>Intrant/Ressource/Moyen</u>	Moyens financiers, humains et matériels utilisés dans un programme ou une politique, par exemple des supports de formation.
<u>Méta-analyse</u>	Processus ou technique ayant pour objet de synthétiser les résultats de recherche d'au moins deux études en utilisant diverses méthodes statistiques afin d'extraire, sélectionner et combiner les résultats d'études précédentes couvrant la même question.
<u>Méta-évaluation</u>	Examen/évaluation de la qualité d'une évaluation. Cette expression est également utilisée pour décrire la synthèse des résultats d'une série d'évaluations afin de juger de leur qualité et/ou d'évaluer la performance des évaluateurs, démarche également appelée synthèse d'évaluation
<u>Méthode quasi expérimentale</u>	Méthode de recherche/d'évaluation qui n'affecte pas de façon aléatoire certaines conditions de traitement aux participants, et dont les groupes témoins sont constitués à l'aide de moyens statistiques. Elle diffère de l'expérimentation (classique) contrôlée, car il n'y a pas de distribution aléatoire du traitement/de l'intervention.
<u>Nettoyage de données</u>	Processus destiné à vérifier le respect des normes, la cohérence interne, l'intégrité référentielle, la validité du domaine des données, et à remplacer/réparer les données erronées par des données correctes. « Nettoyer » un fichier de données signifie vérifier la présence éventuelle de codes erronés et de réponses incohérentes, contrôler que le fichier comporte le nombre correct et attendu d'entrées, de cas et de cartes ou d'entrées par cas et corriger les erreurs trouvées
<u>Résultat/Effet direct</u>	« Quelque chose qui arrive » aux personnes, aux communautés, à l'environnement ou à l'économie, qui est au moins partiellement causé par le programme ou la politique évaluée. Les résultats peuvent être positifs ou négatifs, prévus ou imprévus et concerner des changements qui se produisent ou des changements qui ne pourront plus se produire. Dans le cadre d'une évaluation, l'utilisation du terme « résultat » implique que le programme/la politique a contribué au changement, a participé au changement ou a accéléré le changement
<u>Taille d'effet</u>	Expression statistique désignant l'amplitude de la relation entre les deux variables
<u>Valeur aberrante</u>	Valeur observée dans un ensemble de données qui est très éloignée des autres valeurs de l'ensemble de données. Il s'agit d'une valeur exceptionnellement élevée ou anormalement faible par rapport aux autres.

<p><u>Échantillonnage séquentiel</u></p>	<p>Méthode d'échantillonnage qui peut être aléatoire ou non. En tant que méthode d'échantillonnage aléatoire, l'échantillonnage séquentiel nécessite un cadre ordonné d'unités d'échantillonnage à partir desquelles l'échantillon est sélectionné à l'aide d'une probabilité ou attente explicite, par exemple en sélectionnant chaque dixième cas d'une liste.</p> <p>En tant que méthode d'échantillonnage non aléatoire, l'échantillonnage séquentiel débute par la sélection d'une seule observation ou d'un groupe d'observations qui sont ensuite vérifiées par rapport à l'hypothèse. Si l'hypothèse nulle n'est pas rejetée, une autre observation ou un autre groupe d'observations est échantillonné et l'hypothèse est à nouveau vérifiée, et ainsi de suite.</p>
<p><u>Échantillon aléatoire simple</u></p>	<p>Type particulier d'échantillon aléatoire dans lequel les sujets sont choisis à partir d'une liste à l'aide d'un tableau de nombres aléatoires ou un générateur de nombres aléatoires. (Un échantillon aléatoire nécessite une liste des membres de la population dans laquelle il est possible d'attribuer un numéro particulier à chaque membre.) L'affectation des sujets aux différents traitements, interventions, conditions est réalisée de façon aléatoire plutôt que systématique. La distribution aléatoire des sujets augmente la probabilité que les différences observées entre les groupes résultent de l'intervention expérimentale.</p>
<p><u>Échantillon « boule de neige »</u></p>	<p>Échantillon non aléatoire créé en s'appuyant sur les membres du groupe à étudier pour identifier d'autres membres du groupe (par exemple demander à un sujet à la fin de l'entretien quelle autre personne interroger).</p>
<p><u>Échantillon stratifié</u></p>	<p>Échantillon aléatoire déterminé en divisant la population en groupes ou strates définis par l'existence de certaines caractéristiques, puis par l'échantillonnage aléatoire de chacune des strates. C'est une bonne façon de s'assurer qu'un échantillon d'élèves est diversifié sur le plan racial (par exemple).</p>
<p><u>Revue systématique</u></p>	<p>Méthode de recherche qui fournit un cadre pour réunir et évaluer de façon critique les résultats de données probantes existantes sur une question ou un sujet précis. Les données probantes sont sélectionnées selon des critères précis et la revue est effectuée par le biais d'un protocole normalisé. Après la collecte de toute la documentation pertinente, une évaluation de la qualité des études et rapports est réalisée à l'aide des critères clairement définis et des conclusions sont tirées uniquement à partir des données qui répondent aux normes de qualité établies.</p>
<p><u>Théorie du changement</u></p>	<p>Explique comment les activités s'entendent comme produisant un ensemble de résultats qui contribuent à la réalisation des impacts finaux prévus. Elle peut être élaborée pour tous les niveaux d'intervention : événement, projet, programme, politique, stratégie ou organisation.</p>