



RÉFÉRENTIEL COMMUN

EN ÉVALUATION
DES ACTIONS ET PROGRAMMES
SANTÉ ET SOCIAL

Ce travail a été réalisé par le Groupe évaluation de l'Espace Régional de Santé Publique (ERSP) :

- **Denis FONTAINE**, Médecin de santé publique, Directeur d'études de l'**ORS Rhône-Alpes** (Observatoire Régional de la Santé).
- **Louisa BEYRAGUED**, Sociologue, Chargée de programmes – conseiller méthodologique au **CRAES-CRIPS** (Collège Rhône-Alpes d'Education pour la Santé – Centre Régional d'Information et de Prévention du Sida).
- **Catherine MIACHON**, Ethno-sociologue au **CNDT** (Centre Régional de Prévention des Conduites à Risque) et coordinatrice du **CIRDD** (Centre d'Information et de Ressources sur les Drogues et Dépendances).

Il a été financé dans le cadre du Contrat de Plan Etat – Région par :

- **La DRASS Rhône-Alpes**, Direction Régionale des Affaires Sanitaires et Sociales,
- **La Région Rhône-Alpes**,
- **L'URCAM Rhône-Alpes**, Union Régionale des Caisses d'Assurance Maladie.

ISBN : 2-9521336-1-1

Dépôt légal : mai 2004.

Impression : ORS Rhône-Alpes, 9 quai Jean Moulin, 69001 Lyon

Contact :

Groupe évaluation, Espace Régional de Santé Publique,
9 quai Jean Moulin, 69001 LYON

Tel. : 04 72 10 94 37 – fax : 04 72 07 46 21

Courrier électronique : denis.fontaine@ors-rhone-alpes.org

Site de l'Espace Régional de Santé Publique : www.ersp.org

SOMMAIRE

Introduction : Pourquoi ce document ?	4
--	----------

1^{ère} partie – Les principes de l'évaluation : historique, concepts et limites

Chapitre 1 : Vous avez dit "Evaluation" ?	9
--	----------

- Quatre champs qui convergent vers la nécessité de l'évaluation
- L'évaluation des programmes d'action
- L'évaluation "pragmatique"
- L'évaluation pendant ou à la fin d'un programme

Chapitre 2 : Les principes de l'évaluation	18
---	-----------

- Pourquoi évaluer ?
- L'évaluation, qu'est-ce que c'est ?
- L'évaluation, pour faire quoi ?
- Des dimensions techniques, stratégiques et sociales
- Différencier l'évaluation de ce qu'elle n'est pas

Chapitre 3 : Des enjeux à prendre en compte dans l'organisation de l'évaluation

28

- Des enjeux multiples, explicites et cachés
- Une tension saine et sage
- S'organiser pour susciter la confiance
- Attention aux objectifs parasites : ni audit, ni publicité
- Qui fait quoi ?

Chapitre 4 : Evaluer quoi ? Les objets de l'évaluation

40

- Les cinq objets d'évaluation d'un programme
- Evaluer les efforts
- Evaluer les effets

2ème partie – La pratique de l'évaluation : méthodes et étapes

Chapitre 5 : Comment évaluer ? La méthodologie 54

- Modèle axé vers les buts ou modèle systémique ?
- L'approche systémique du processus et de l'impact
- L'approche épidémiologique des effets
- Les difficultés de l'évaluation : mesure et appréciation

Chapitre 6 : Les étapes d'une évaluation 70

- Le préalable : l'évaluation est-elle pertinente et réalisable ?
- La préparation : du comité d'évaluation au cahier des charges
- La réalisation : du recueil des données au rapport final
- La valorisation : diffusion des résultats et suivi post-évaluation

Conclusion : pour un langage commun en évaluation 83

Annexes :

1. Lexique de l'évaluation et index 87
2. Bibliographie 97

Introduction : Pourquoi ce document ?

"Il faut évaluer !". L'injonction est forte. Elle vient d'en haut : les décideurs le martèlent, les financeurs l'exigent. Elle vient aussi d'en bas : le bénéficiaire est souvent aussi un contribuable, qui parfois doute de l'utilité de certaines actions. Quant aux acteurs des programmes de santé ou du social, souvent débordés de travail, ils s'interrogent sur leurs résultats... ou regrettent qu'ils ne soient pas mieux reconnus.

L'évaluation déclenche aussi des **réticences**, voire *une certaine résistance instinctive* [34], pour deux raisons principales. D'une part, elle *est parfois envisagée comme un moyen de contrôle de la part des financeurs, des décideurs ou du pouvoir de tutelle en général* [2]. D'autre part, elle a un coût¹ qui peut paraître insupportable dans un contexte de ressources rares et de calendriers de travail surchargés.

Ces réticences sont fondées sur des inquiétudes légitimes, car il y a d'une part des enjeux politiques, donc des risques de manipulation dans toute évaluation, et d'autre part il faut des moyens humains et financiers pour évaluer une action. Mais elles peuvent aussi cacher des **difficultés** tenant au sujet même de l'évaluation :

- Il est difficile de se comprendre quand on parle d'évaluation. Tout le monde n'emploie pas les mêmes termes pour désigner les mêmes choses. Ou à l'inverse, tout le monde ne met pas le même sens derrière les mêmes mots.
- Il existe tellement d'éléments à évaluer dans une action ou un programme relevant du domaine de la santé ou du social que parfois on ne sait pas sur quoi l'évaluation doit porter précisément pour être opportune.

Partant du principe qu'il est préférable de s'emparer de la question de l'évaluation pour en faire un outil de travail, plutôt que de se la voir imposée, nous avons réalisé ce document. Il a pour **objectif** de proposer un **référentiel commun pour l'évaluation des programmes** santé ou social,

¹ "On estime généralement que les moyens financiers à dégager pour une évaluation représentent entre 5% et 20% des dépenses globales d'un programme de santé" [2]. Toutefois, on ne peut en tirer une règle. Plus le programme est petit, plus la part de l'évaluation sera élevée.

c'est-à-dire un langage, des concepts et des méthodes partagés par l'Espace Régional de Santé Publique Rhône-Alpes et ses partenaires.

Nous avons donc tenté de répondre aux trois principales questions que se pose toute personne concernée par l'évaluation d'un programme santé ou social :

1. De quoi parle-t-on au juste ?
2. Quel(s) type(s) d'évaluation est(sont) utile(s) pour le programme tel que nous le connaissons aujourd'hui ?
3. Comment nous y prendre pour réaliser cette évaluation ?

Pour cela, nous avons relu les documents méthodologiques de l'évaluation et tenté d'extraire leurs éléments essentiels, communs et éventuellement divergents. Pour plus de "robustesse" dans l'analyse, **quatre domaines d'application** de l'évaluation ayant des références différentes ont été étudiés : la santé (les soins et les programmes de santé), l'éducation pour la santé / promotion de la santé, le domaine social (incluant le développement social urbain) et l'aide aux pays en développement.

Chaque domaine n'a pas été exploré exhaustivement, car la littérature consacrée à l'évaluation est d'un volume très important. Nous avons fait une revue sélective des documents les plus importants publiés en langue française² des années 90 à 2001, soit qu'ils fassent autorité, soit qu'ils soient très pédagogiques.

L'originalité de ce travail est de s'appuyer sur de nombreuses citations extraites des 37 publications analysées, afin de rester aussi proche que possible de ces sources. Pour faciliter la lecture, *les citations sont simplement notées en italiques, sans utiliser les guillemets.*

Toutefois, ce référentiel n'est pas une simple compilation de la bibliographie. Il constitue **une synthèse de toutes les notions clés nécessaires pour évaluer** des programmes dans le domaine de la santé ou du secteur social.

Enfin, il est orienté vers la pratique de l'évaluation et non vers la seule connaissance. Il reflète donc **les choix que nous avons faits**, en raison de la pléthore de vocabulaire et des divergences de point de vue qui existent parfois entre les auteurs. Ces principaux choix sont les suivants :

² Les auteurs américains classiques ont pratiquement tous été traduits et adaptés en français, notamment par les Québécois.

- L'évaluation est le moteur du changement, plus que la phase de planification. Les acteurs sont souvent pris dans une logique d'action qui les pousse à initier leurs programmes à partir d'un diagnostic souvent implicite et intuitif, basé sur leur connaissance du "terrain". Ce n'est qu'après un certain temps que les questions émergent, face à des problèmes, ou quand se pose la question des effets du programme. C'est pourquoi l'évaluation est envisagée dans ce document pour les programmes ayant déjà débuté. Bien entendu, il est préférable de prévoir l'évaluation dès la planification du programme.
- Les acteurs des programmes sont les premiers acteurs de l'évaluation. C'est par eux que passe le développement d'une "culture d'évaluation". Même externe, une évaluation a peu de chances de succès si elle n'obtient pas l'adhésion des acteurs du programme. Ce document se place d'abord de leur point de vue, sachant qu'une évaluation doit cependant prendre en compte les regards et les attentes des différents partenaires (financeurs, bénéficiaires, etc.). C'est pourquoi le modèle de l'évaluation participative et négociée est privilégié.
- L'évaluation n'est jamais un exercice banal, et les réticences qu'elle entraîne ne doivent pas être sous-estimées. Toute évaluation "systématique" et obligatoire court le risque d'être neutralisée et de devenir un exercice de style anecdotique. L'évaluation doit donc s'inscrire à un moment pertinent de la vie du programme et répondre à un réel besoin de connaissance en vue d'une prise de décision.
- L'évaluation n'est jamais un exercice simple, car un programme repose sur les interactions complexes entre plusieurs acteurs. Il n'y a de bonne évaluation que "sur-mesure". La méthodologie ne peut donc être simplifiée et standardisée. De même le vocabulaire, qui reflète des histoires professionnelles et des regards divers. Nous avons pris le parti de rassembler les concepts en grande catégories, quitte à prendre pour synonymes des mots qui ne se recouvrent pas exactement pour les spécialistes.

Ce document comporte trois parties :

1. Dans la **1^{ère} partie** (chapitres 1 à 4) sont abordés les **concepts** principaux de l'évaluation : ses définitions, ses logiques, ses enjeux, ses objets. Cette première partie correspond à la phase de planification de l'évaluation.
2. Dans la **2^{ème} partie** (chapitres 4 et 5) sont abordés les **aspects pratiques** de l'évaluation : les méthodes, les étapes, les facteurs de réussite (les critères de qualité) de l'évaluation. Cette seconde partie correspond à la mise en œuvre de l'évaluation.

3. Enfin, en **annexes**, un **lexique** reprend les définitions des principaux termes utilisés dans ce document, et pour en savoir plus, les références bibliographiques sont détaillées.

Chaque partie peut être lue de façon indépendante.

Pour une lecture rapide et plus opératoire, vous pouvez lire les encadrés de début de chapitre ("L'essentiel"), ainsi que la conclusion.

Si vous abordez l'évaluation pour la première fois, nous vous suggérons de commencer par notre document "l'évaluation en neuf étapes"³.

³ Document téléchargeable sur le site de l'ERSP : www.ersp.org

1^{ERE} PARTIE :

LES PRINCIPES DE L'EVALUATION :

HISTORIQUE, CONCEPTS ET LIMITES

Chapitre 1 : Vous avez dit "Evaluation" ?

L'essentiel :

- L'évaluation apparaît de plus en plus comme un exercice incontournable dans les secteurs de la santé et du social. Cependant, il s'agit d'un domaine encore récent.
- Située à la jonction de la connaissance et de l'action, l'évaluation ne repose ni sur un cadre conceptuel unique, ni sur un vocabulaire standardisé. Tout exercice d'évaluation doit donc s'accompagner d'une clarification de l'approche et des mots utilisés.
- Ce référentiel est basé sur l'analyse de la bibliographie dans quatre champs de l'évaluation : la santé (les soins et programmes), l'éducation pour la santé, le domaine social (notamment le développement social urbain) et l'aide aux pays en développement.
- Il traite de l'évaluation d'un programme ou d'une action mis en oeuvre dans le but de résoudre un ou des problèmes dans le domaine de la santé ou du social. Il s'intéresse à l'évaluation qui s'effectue en cours ou à la fin du programme ou de l'action. Le diagnostic avant le lancement d'un programme n'est pas abordé ici.
- Il faut différencier l'évaluation d'un programme, qui s'intéresse à des actions, et la supervision, qui s'intéresse aux personnes, cette dernière s'inscrivant dans un cadre de management.
- Ce document ne traite pas de la supervision des personnes, de l'évaluation des outils, de l'évaluation des politiques, ni d'évaluation économique.

L'évaluation peut s'appliquer dans de multiples domaines et à de nombreux objets. Nous avons donc dû cibler les champs d'application de l'évaluation. Ce chapitre comporte un rapide **historique** des démarches d'évaluation, puis il trace les **limites d'application des concepts** d'évaluation développés dans ce document.

Historiquement, les premières demandes d'évaluation se situent dans le champ de l'éducation et remontent au début du 20^{ème} siècle. Elles portent sur la mesure de la performance, en vue d'une recherche d'efficacité des systèmes éducatifs. Les pouvoirs publics *doivent faire face ainsi à des questions inédites et cherchent du côté de la science des réponses que les systèmes politique et administratif du 19^{ème} siècle sont impuissants à fournir* [26],[6].

C'est ensuite à partir de la question de l'allocation optimale des ressources que se développe l'évaluation, que ce soit dans le champ de l'action publique ou dans celui de l'entreprise. Ces développements nord-américains de la planification par projet sont repris en France après la seconde guerre mondiale. *La planification économique et sociale a sans doute contribué [depuis les années soixante-dix] à instaurer et développer les démarches d'évaluation de l'efficacité des politiques, y compris celles du domaine sanitaire et social : Rationalisation des Choix Budgétaires (RCB), mesure coût-efficacité, audits et études de productivité, missions des grands corps de contrôle de l'état, observatoires divers, etc. L'objectif le plus souvent recherché est d'améliorer la qualité du service rendu à l'usager. Il est aussi, dorénavant, d'optimiser les dépenses... Les politiques contractuelles se sont d'ailleurs développées à tous les niveaux de l'action sociale : les financeurs imposent une claire définition des objectifs, moyens et stratégies, et aussi un dispositif d'évaluation des résultats* [10].

Il ne faudrait cependant pas analyser la montée des pratiques d'évaluation sur le mode d'un véritable raz-de-marée. Au contraire, son développement fut assez lent [...] en dépit de quelques programmes phares. C'est surtout à partir de 1980 que la mécanique s'emballe. Une nouvelle étape commence en 1990 avec la mise en place [...] d'un dispositif assez complexe d'impulsion et de régulation de l'évaluation [les politiques publiques] : Comité Inter-Ministériel de l'Evaluation [CIME], fonds de financement et Conseil Supérieur de l'Evaluation [CSE] [10].

Quatre champs qui convergent vers la nécessité de l'évaluation

Venant d'horizons variés, et étant amenés à évaluer des programmes très différents dans le domaine sanitaire et social, nous avons ciblé notre

recherche documentaire sur **quatre domaines** d'application de l'évaluation, qui ont chacun leur histoire : la santé (les soins...), l'éducation pour la santé, le domaine social (et notamment le développement social urbain), et l'aide aux pays en développement.

① Dans le **domaine de la santé**, les auteurs américains sont les premiers à avoir étudié l'évaluation des pratiques médicales et de la qualité des soins dans les années soixante et soixante-dix. Parallèlement, s'imposait la nécessité de soumettre les nouvelles thérapeutiques à une évaluation de leur efficacité avant leur diffusion. Dans les deux cas il y avait l'idée de rationaliser les choix, de les appuyer sur des arguments scientifiques, dans un double objectif d'efficacité et de minimisation des coûts. Les Québécois ont ensuite repris ces travaux dans les années quatre-vingts et les ont adaptés de façon plus large aux programmes de santé, incluant la prévention et l'approche communautaire, et non plus seulement les soins curatifs. L'ouvrage de R. Pineault et C. Daveluy est ainsi une référence dans ce domaine [35].

En France, l'Andem (Agence nationale pour le développement de l'évaluation médicale), devenue depuis l'ANAES (Agence nationale de l'accréditation et de l'évaluation en santé) a repris ces travaux et elle a produit un guide pour l'évaluation d'une action de santé publique, qui est le principal document méthodologique français [1].

② En **éducation pour la santé**, l'évaluation s'est principalement développée dans les années quatre-vingts, en s'inspirant des méthodes de la santé publique.

Dans le champ de l'éducation pour la santé on peut distinguer deux formes : [l'une qu'on peut qualifier] de *profane qui est le fait de tout éducateur (ex. : parent)* [et l'autre] dite "*savante*" qui repose sur un corpus de théories explicatives. [La première] est permanente et s'inscrit dans les gestes quotidiens. Son efficacité est difficilement quantifiable [mais] son impact est certainement prépondérant. [La seconde] est le fait de professionnels et est censée utiliser un répertoire d'interventions dont les effets ont fait l'objet d'une évaluation. Dans les deux cas il s'agit de favoriser l'adoption d'attitudes⁴ et de comportements conduisant à une meilleure santé pour l'individu ou la population, à court ou à long terme [26].

Au cours de ces dernières décennies, de nombreux programmes en éducation pour la santé et en prévention des comportements à risque ont été développés. Les évaluations de ces programmes d'éducation pour la

⁴ Une attitude est une disposition polarisée, pour ou contre quelqu'un ou quelque chose, qui pousse à un comportement. S'agissant d'un état d'esprit propre à la personne, elle n'est pas observable mais elle peut être étudiée à partir de ses déclarations.

santé ont montré des effets limités en termes de changement de comportements. Ce constat a conduit à une réflexion d'ensemble et à un développement de l'évaluation. *L'amélioration des méthodes d'intervention éducative et des techniques d'évaluation a surtout été le fait d'équipes de recherche nord-américaines, scandinaves ou hollandaises. [...] Dans ces pays, il existe en général des pôles universitaires de psychologie sociale cognitive faisant suite à l'école de psychologie comportementale et des pôles universitaires de santé publique intéressés par les comportements de santé* [26].

En matière d'évaluation, deux concepts d'interventions coexistent et se justifient chacun par le type d'approche utilisé dans les programmes. Les tenants d'une première école se réfèrent à l'approche "épidémiologiste" [...]. Ceux d'une deuxième école, se réfèrent aux approches "communautaires". [...] La première école d'évaluateurs utilise la méthode expérimentale qui consiste à comparer l'évolution d'une population ayant bénéficié d'un programme éducatif à celle d'une population n'en ayant pas bénéficié ou ayant bénéficié d'un programme alternatif. La seconde école d'évaluateurs propose de mettre l'accent sur l'évaluation de processus, c'est-à-dire du pilotage fin des actions réalisées, en particulier par des évaluations qualitatives permettant non pas de mesurer un changement mais de comprendre comment le programme a été appliqué, compris, vécu, accepté. [...] La première école défendrait donc une efficacité scientifiquement déterminée à partir d'indicateurs quantitatifs, et la seconde une efficacité démocratiquement déterminée, c'est-à-dire par un jury comprenant financeurs, population, intervenants et experts [26]. Actuellement, il est intéressant de noter que les tenants des deux écoles d'évaluateurs ne s'opposent pas et se rejoignent sur différents points.

Enfin, en éducation pour la santé, les campagnes nationale de communication grand public sont évaluées, mais *on ne mesure que très moyennement la qualité d'une communication. Lorsqu'on parle d'évaluation de campagne, on parle le plus souvent de post-test. Les post-tests portent sur un échantillon représentatif de la population ciblée par la campagne et mesurent au minimum : l'impact, la mémorisation, l'appréciation, le pouvoir d'incitation et le pouvoir d'implication*⁵.

Pour M. Demarteau, *le développement de l'évaluation est lié au développement de l'éducation pour la santé. Il faut développer de nouveaux modes d'action en même temps que de nouveaux modes d'évaluation* [16].

⁵ Chauvin F., Brixi O., Roussille B. (Eds). **Du bon usage de la communication en éducation pour la santé. Séminaire international Château de Bierville (Essonne), 16 et 17 décembre 1996.** Vanves : Editions CFES, Avril 1998, 289 p.

③ Dans le **domaine social**, les auteurs américains ont réfléchi dès les années trente à l'évaluation des programmes, mais c'est dans les années soixante et soixante dix que la réflexion sur l'évaluation se structure.

En France, l'évaluation émerge à partir d'un double mouvement : répondre à une demande sociale d'amélioration des conditions de vie dans les quartiers d'habitat social souvent situés en périphérie des grandes villes et évaluer les réponses apportées dans le cadre des dispositifs de la **politique de la ville** (Habitat et Vie Sociale, Développement Social des Quartiers, Développement Social Urbain, Contrat de Ville). Il s'agit de voir en quoi ce type de politique publique conduit à un impact sur les phénomènes sociaux, particulièrement à la suite des violences urbaines des années quatre-vingts, qui ont vu une population de jeunes résidant en "banlieue" manifester pour leur intégration dans les activités sociales et professionnelles de la société française. Les programmes développés dans le cadre de la politique de la ville favorisent l'émergence de nouveaux métiers et de nouveaux acteurs, interface entre les populations et les institutions. Dans ce cadre, l'évaluation devient un outil d'apprentissage et de mobilisation pour les politiques publiques visant à provoquer des changements sociaux puis à adapter les orientations politiques.

L'évaluation est liée à la fois à la recherche, en s'appuyant sur l'analyse de l'évolution des problématiques locales et l'expérimentation sur le changement social, et aux orientations des politiques publiques, avec la création de la Délégation Interministérielle à la Ville. Elle doit contribuer à produire du débat public, en favorisant la notion de "citoyenneté" principalement recherchée à partir de la "participation des habitants" à la politique publique.

Au fil des années, les évaluations ont abandonné une logique trop simpliste visant à relier des objectifs, des moyens et des résultats au profit d'une logique s'efforçant de rapprocher les acteurs, les problèmes, les interventions et les effets. Cette démarche pour appréhender la complexité du social a eu son revers en terme de difficulté à maîtriser les processus de production de l'évaluation. [Toutefois], l'évaluation n'est pas l'outil définitif de régulation du social, elle est [...] un outil parmi d'autres, la contribution au développement d'une culture [plus] qu'une aide déterminante à la prise de décision [10].

④ Dans **les pays en développement**, la rareté des ressources rend plus forte la contrainte des choix de priorités et de la planification, et par conséquent l'évaluation. Dans la cadre de la généralisation de la stratégie des soins de santé primaires lancée en 1978, l'Organisation Mondiale de la Santé a apporté sa contribution à la diffusion des connaissances sur l'évaluation en publiant en 1981 des *principes directeurs* pour l'évaluation des programmes de santé [34]. Les pays donateurs de l'Aide Publique au

Développement ont également encouragé l'évaluation dans les années quatre-vingts, dans tous les secteurs, dont celui de la santé.

En France, le Ministère de la Coopération⁶ a repris la préoccupation générale d'évaluation des politiques publiques et il a publié un guide des principes de l'évaluation [31]. La préoccupation d'évaluation touche également les Organisations Non Gouvernementales (ONG), pour qui la recherche de l'efficacité passe plus par des actions de proximité que par de grands programmes nationaux. Mais la demande des bailleurs de fonds a là aussi été un facteur de développement de l'évaluation. Dans les années quatre-vingt-dix, un fonds a été constitué pour accompagner les ONG dans leurs démarches d'évaluation, des débats et formations ont été organisés sur ce sujet. La Fondation de France, qui finance des actions de développement local, a également publié un document synthétique sur l'évaluation des actions de développement [19].

L'évaluation des programmes d'action

Un auteur américain (Patton), a recensé dès 1981 quelques 132 formes d'évaluation. Il faut donc préciser ce qui nous intéresse dans ce document. Nous nous sommes limités ici à l'évaluation d'**actions**, d'**interventions**, de **programmes** et/ou de **projets**. Tous les auteurs n'entendent pas ces mots comme équivalents : les programmes ou projets sont des ensembles coordonnés d'actions ou d'interventions. Nous emploierons dans ce document le terme général de "programme" pour désigner l'un ou l'autre de ces termes, car la démarche d'évaluation est la même, quel que soit le degré de complexité d'une intervention (Cf. encadré 1).

Les **réseaux** (de soins, de prévention, ou médico-sociaux) se rapprochent beaucoup des programmes par leur aspect partenarial et peuvent être évalués comme eux⁷.

Les "programmes" ainsi définis se différencient d'une part des "techniques" et des "pratiques", et d'autre part des "politiques", dont l'évaluation n'est pas abordée dans ce document :

- Les **outils et techniques** (que ce soient des techniques de soin, d'animation...) et les **pratiques** (les actes accomplis par les

⁶ avant son intégration dans le Ministère des Affaires Etrangères.

⁷ Pour plus de détails sur l'évaluation des réseaux, voir les publications de l'Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES) :

- **Principes d'évaluation des réseaux de santé**. Paris : ANAES, 1999, 140 p.
- **Evaluation des réseaux de soins. Bilan de l'existant et cadre méthodologique**. Paris : ANAES, 2001, 63 p. Publication en ligne sur le site www.anaes.fr.

professionnels dans le cadre de leurs activités), sont d'un niveau "inférieur". Ce sont les "briques" des programmes, puisque les actions (dont l'ensemble forme un programme) sont basées sur des outils et des techniques (qui peuvent concerner les soins, les équipements, l'animation, la formation...).

- Les **politiques** sont d'un niveau "supérieur", puisqu'elles se composent d'actions de type législatif, réglementaire, de l'organisation des ressources humaines et matérielles, etc., en plus de programmes. Mais les programmes étant l'une des composantes d'une politique, l'évaluation d'un programme contribue en partie à l'évaluation de la politique dans laquelle il s'insère.

Encadré 1. Définition du programme.

Un programme est *l'ensemble des moyens (physiques, humains, financiers [...]) organisés dans un contexte spécifique, à un moment donné, pour produire des biens ou des services dans le but de modifier une situation problématique* [11]. La notion de programme s'applique [...] aussi bien au domaine de la prévention et de la promotion de la santé qu'à celui des services [de soins ou sociaux] [17].

Il existe d'autres définitions du programme. Toutes décrivent les mêmes éléments constitutifs : des ressources organisées, des activités ou des services réalisés par des acteurs (et en général destinées à des groupes cibles ou bénéficiaires), des objectifs (liés à des situations problématiques), un contexte donné, des résultats et des conséquences ou impact [12]. Ces éléments seront aussi les principaux points d'entrée de l'évaluation (Cf. chapitre 4).

Par exemple :

- Un ordinateur qui simule la courbe d'alcoolémie en fonction de la quantité bu est un outil de prévention.
- Ce logiciel peut être utilisé lors d'animations auprès des clients des discothèques, qui constituent une pratique de prévention.
- Un programme de prévention de la conduite en état d'alcoolémie peut être constitué de plusieurs actions : des animations auprès des clients des discothèques, des autorisations de fermeture plus tardive des discothèques sans distribution d'alcool, la promotion des "capitaines de soirée" (conducteur sobre), etc.

- Une politique de prévention de l'alcoolisation excessive s'appuie sur plusieurs programmes, tels que le précédent, mais aussi sur des mesures comme la répression de l'alcoolisation au volant, la limitation de la publicité pour l'alcool, le développement de consultations d'alcoologie, etc.

Enfin, il faut différencier l'évaluation des programmes de **la supervision des personnes**. L'évaluation a pour objet un programme, qui est analysé en vue d'une décision le concernant. On n'évalue pas les personnes, on les supervise. La supervision est tournée vers les personnes. Elle fait partie du management des ressources humaines et a pour but l'amélioration (ou le maintien) de la performance⁸ des personnes.

L'évaluation "pragmatique"

La plupart des auteurs opposent l'évaluation pragmatique et l'évaluation de recherche (ou recherche évaluative) [1],[3],[11],[24] :

- **L'évaluation pragmatique**⁹ est réalisée par des professionnels de terrain dans le cadre de leurs pratiques. Elle *a pour but de comparer la pratique quotidienne aux données acquises de la science puis d'examiner les causes d'une éventuelle discordance afin d'y remédier* [24]. Elle vise à s'assurer de la conformité de la mise en œuvre du programme et/ou à l'obtention des effets attendus, donc *l'amélioration de la performance* du programme [3].
- **L'évaluation de recherche** est pratiquée par des chercheurs et *a pour finalité de définir les pratiques optimales, autrement dit elle contribue à remettre en cause et à améliorer les données de la science, qui ne sont jamais définitivement acquise* [24]. Elle vise à établir l'efficacité de nouvelles interventions, autrement dit elle a une *finalité fondamentale* [12], qui est *la poursuite de la connaissance* [3]. Un exemple de ce type d'évaluation est l'essai thérapeutique, où l'efficacité d'un nouveau traitement est comparée à celle d'un ancien traitement.

D'autres auteurs définissent l'évaluation de recherche de façon plus large. Pour eux, il y a recherche évaluative dès lors qu'on met en place une démarche scientifique pour tester la relation entre une intervention et ses effets. Suivant cette définition, l'évaluation de recherche concerne également les programmes professionnels. Les méthodes diffèrent peu

⁸ Pour plus de détails sur la supervision, voir par exemple : Flahault D., Piot M., Franklin A. **La supervision des personnels de santé au niveau du district**. Genève : OMS (Organisation Mondiale de la Santé), 1988, 107 p.

⁹ Ou évaluation de routine, évaluation professionnelle

entre ces deux types d'évaluation. Leur distinction a donc peu d'intérêt pratique.

L'évaluation pendant ou à la fin d'un programme

L'analyse ou diagnostic de situation¹⁰ est une forme particulière d'évaluation qui se fait avant de lancer une action. Elle a une *finalité stratégique* [12], elle permet de valider la planification avant de lancer le programme. Nous l'avons exclue de ce travail car il nous paraît plus conforme aux attentes des acteurs de réfléchir aux programmes déjà en place. D'autre part, elle est tellement liée à la planification qu'elle ne peut en être dissociée.

Nous ne nous intéresserons ici qu'à l'évaluation concernant des **actions déjà mises en place ou qui se terminent** : c'est l'évaluation *post-décision* [35] ou encore *évaluation rétrospective* [31], *qui fournit des informations sur la mise en place, le fonctionnement et les effets du programme* [35].

Toutefois, cette évaluation post-décision doit **être prévue le plus tôt possible**, c'est-à-dire dès la conception du programme. En effet, plus le diagnostic aura été précis, plus les objectifs et le plan d'action auront été définis avec précision, plus l'évaluation sera facile à organiser. Néanmoins, si le programme a débuté sans que ces éléments aient été formalisés, il sera possible de les reconstituer a posteriori pour construire l'évaluation.

Enfin, **l'évaluation économique**¹¹ des programmes n'est pas non plus détaillée dans ce document. Elle est basée sur l'analyse des coûts des ressources mobilisées pour un programme ou un service. Ces coûts sont ensuite mis en relation avec des activités produites (analyse de la productivité), avec des résultats obtenus (analyse coût-utilité, analyse coût-efficacité ou efficacité¹² ou rendement), ou encore avec toutes les conséquences possibles (analyse coût-bénéfice). L'évaluation économique n'est pas abordée ici en raison de ses méthodes spécifiques et parce qu'elle ne correspond pas aux questions fréquemment posées lors d'une évaluation de terrain.

¹⁰ Elle porte de nombreux autres noms : état des lieux, analyse ou évaluation des besoins, étude de milieu, analyse préliminaire, évaluation prospective, évaluation de faisabilité, évaluation de la planification, évaluation *a priori*, évaluation *ex-ante* ou pré-décision.

¹¹ Pour plus d'informations sur l'évaluation économique, voir par exemple :

- Drummond M.F., O'Brien B., Stoddart G., Torrance G.W. **Méthodes d'évaluation économique des programmes de santé**. Paris : Economica, 1998, 331 p.
- Beresniak A., Duru G. **Economie de la santé**. Paris : Masson, 1992, 157 p.

¹² L'efficacité est brièvement abordée page 50.

Chapitre 2 : Les principes de l'évaluation

L'essentiel :

- L'évaluation est un outil de mesure et de conduite du changement. Elle a deux buts principaux : le but formatif, qui vise à améliorer la mise en œuvre du programme (le pilotage) ; le but sommatif, qui analyse les effets du programme pour en tirer les leçons.
- Evaluer, c'est répondre à la question générique "Est-ce que ça marche ?", en vue d'une décision concernant le programme.
- Toutes les définitions de l'évaluation associent trois dimensions : 1/ Une collecte d'informations systématique et diversifiée sur le programme. 2/ Une appréciation critique portée en référence à des critères et des normes. 3/ Des recommandations en vue d'une décision.
- On ne peut pas ne pas évaluer. Toute personne, toute organisation a des pratiques d'évaluation intuitives. Elles ne sont pas suffisantes mais peuvent servir de base pour construire une évaluation fondée sur une analyse systématique du programme, aussi explicite et formalisée que possible.
- On ne peut pas tout évaluer. Il faut donc préciser les questions les plus importantes qu'on se pose sur le programme, et l'utilisation qui sera faite des résultats de l'évaluation (les décisions à instruire). Autrement dit, répondre aux questions : Pourquoi évaluer ? Evaluer pour faire quoi ?

Ce document s'intéresse, rappelons-le, à l'évaluation qui s'effectue en cours ou à la fin d'un programme d'action mené par des professionnels de terrain dans le but de résoudre un ou des problèmes dans le domaine de la santé ou du social.

Il faut évaluer ! Mais pourquoi ? Les questions qui sont donc posées à l'évaluation, sont souvent les mêmes : "Est-ce que ça marche ?" ; "Est-ce que ça sert à quelque chose ?" ; "Qu'est-ce qui a changé ?".

Pourquoi évaluer ?

L'évaluation se place dans le *cycle de la planification* [4]. Classiquement, *l'évaluation ferme la boucle de la planification* (Cf. Schéma 1) [35]. Cependant, il faut en avoir une vision plus large : elle *doit être considérée comme une partie intégrante et continue du travail d'ensemble de la planification : elle est un processus d'analyse présent à chaque étape comme en fin d'action.* [7] ; dans cette conception, l'évaluation est au centre du processus de planification (Cf. Schéma 2).

Schéma 1 : La place de l'évaluation dans le cycle de planification

Source : Pineault R. et Daveluy C., 1995 [35]

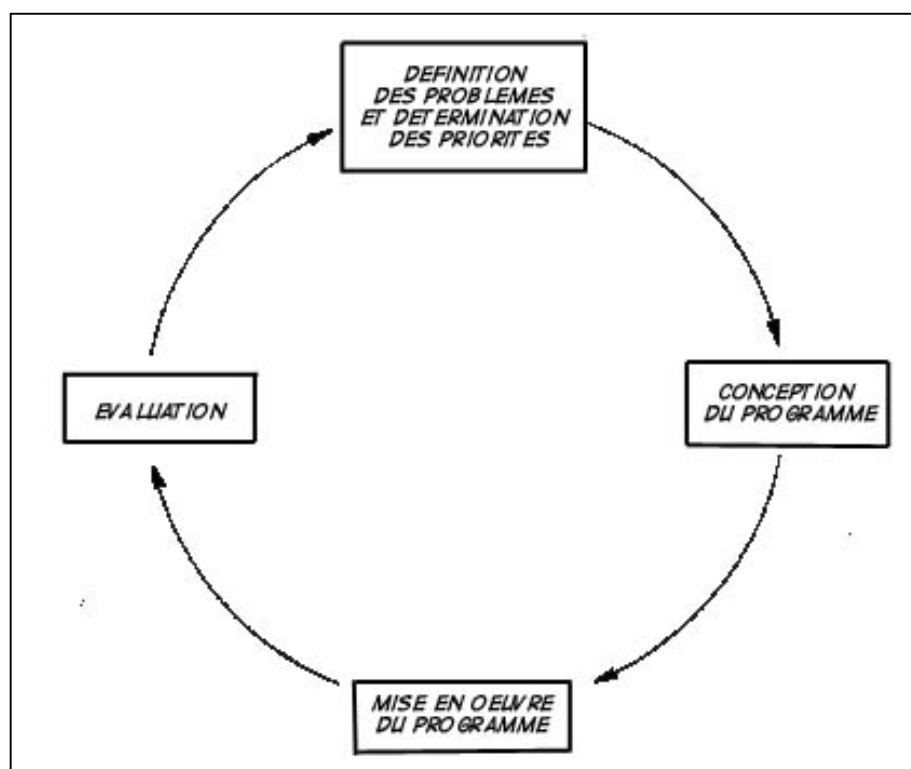
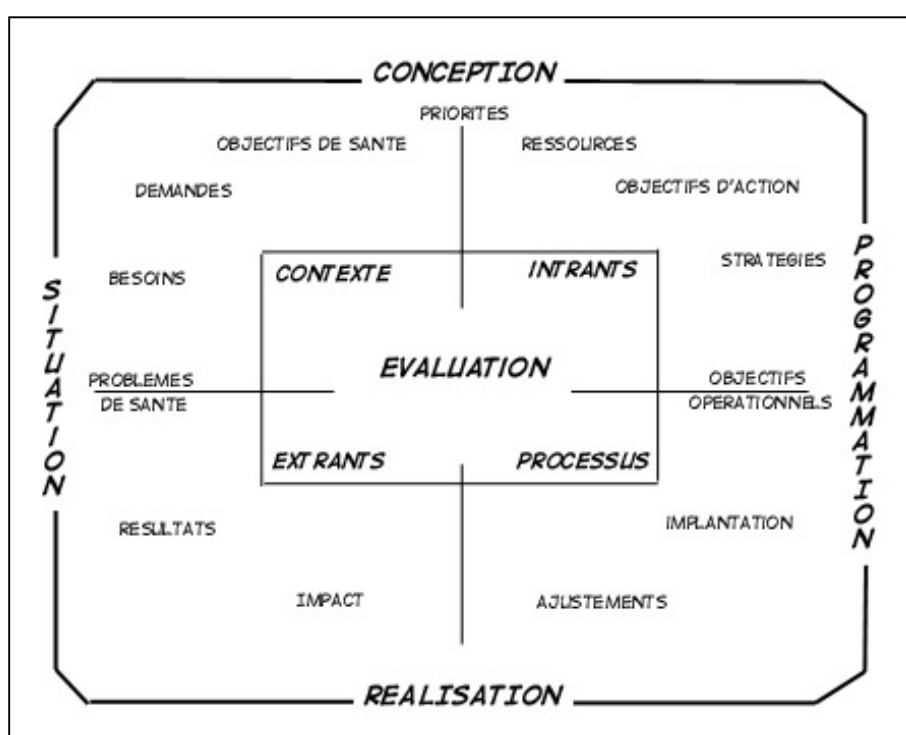


Schéma 2 : La place de l'évaluation dans la planification d'intervention d'éducation pour la santé
Source : Bury J.A., 1988 [7]



Pourquoi évaluer un programme ? C'est la première question à se poser avant d'organiser une évaluation. Les auteurs distinguent deux grands **buts** d'évaluation : le but **formatif** (en cours de programme) et le but **sommatif** (en fin de programme).

1. **L'évaluation formative** maintient une rétroaction constante tout au long du processus d'élaboration et d'implantation du programme, permettant à celui-ci d'être continuellement modifié à mesure que des données nouvelles sont connues [35]. Il s'agit de fournir de l'information pour l'amélioration, la modification ou en général la gestion du programme [9]. Elle favorise une meilleure connaissance de l'action, permet une prise de conscience critique du travail réalisé, fournit des repères pour réajuster le tir et maintenir le cap : elle est apprentissage dans l'action et action autocorrectrice [30].

Il s'agit d'évaluation concurrente ou concomitante qui survient en même temps que les étapes du programme se déroulent [35], ou évaluation en cours de programme¹³ [1].

2. ***L'évaluation sommative¹⁴*** s'attache aux résultats obtenus par une intervention ou un programme [35]. Plus généralement, elle s'intéresse à toutes les conséquences du programme : il s'agit de *déterminer les résultats et l'impact du programme* [9]. Cette évaluation permet de *comprendre les effets induits dans leur globalité*, et ce lien entre action et effet permet de *donner un sens aux actions* [14].

L'ANAES (ex ANDEM) distingue deux formes d'évaluations sommatives : *les évaluations strictement comparatives de type avant/après ou ici/ailleurs* qui permettent de montrer l'efficacité du programme par rapport à un groupe témoin qui n'en a pas bénéficié, et *les évaluations analytiques qui recherchent les causes de dysfonctionnement ou de changement* [1].

L'évaluation sommative se situe à la fin du programme ou d'une de ses phases. Selon le cas, il s'agit de *l'évaluation intermédiaire* (en fin de phase) ou de *l'évaluation finale* [1]. Elle peut aussi se situer un certain temps après la fin du programme : *l'évaluation ex-post¹⁵* permet de *vérifier l'impact ou le prolongement de l'impact du projet* et d'en apprécier la durabilité [2].

Les objectifs de l'évaluation se traduisent en questions posées par rapport au programme. La question générique de toute évaluation ("Est-ce que ça marche ?") peut se décliner en **grandes questions** [35] :

1. *Est-ce que la mise en œuvre du programme s'est faite telle que prévu ?*
2. *Est-ce que les ressources mobilisées étaient adéquates ?*
3. *Est-ce que les actions mises en place étaient appropriées ?*
4. *Est-ce que les objectifs spécifiques du programme ont été atteints ?*
5. *Est-ce que le programme tel que conçu et réalisé correspond à des priorités ?*
6. *Est-ce qu'il répond à des problèmes importants du point de vue de la société et [est-ce] qu'il va dans la direction attendue ? [35].*

¹³ Autres termes pour l'évaluation formative : *évaluation continue* [13], *évaluation permanente* [7],[8], *évaluation endoformative* pour renforcer l'idée que cette évaluation est souvent faite par les acteurs eux-mêmes et pour eux-mêmes [29], ou encore *évaluation dynamique de type rétro-action* [1].

¹⁴ Autres termes pour l'évaluation sommative : *évaluation récapitulative* [29], *évaluation globale* [7],[8], *évaluation-bilan* [37].

¹⁵ Tous les auteurs ne font pas cette distinction : beaucoup considèrent comme synonymes évaluation ex-post et évaluation à la fin du programme.

L'évaluation, qu'est-ce que c'est ?

En raison de la diversité des regards sur l'évaluation, il n'existe pas de définition unique de l'évaluation de programme. L'encadré 2 donne quelques définitions synthétiques qui montrent le sens de l'évaluation.

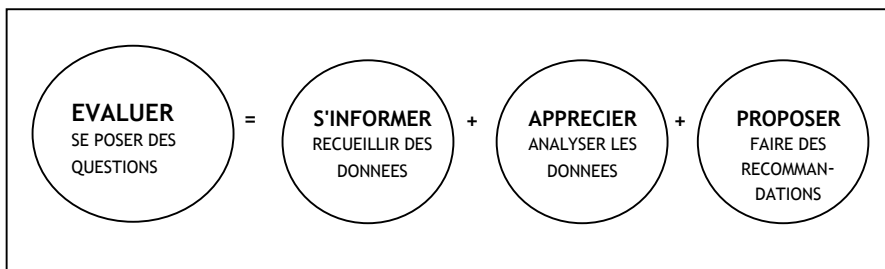
Encadré 2. Quelques définitions courtes de l'évaluation de programme.

De façon courte et concrète, l'évaluation c'est :

- *Une réflexion sur l'action* [16].
- *Un moyen systématique de tirer les leçons de l'expérience* [34].
- *Mieux comprendre afin de mieux agir* [3].
- *Aider à révéler le besoin, à révéler la nature des logiques et des contradictions, faire de la maïeutique, de l'entremêlage* [27].
- *Se poser des questions et se donner les moyens d'y répondre* [1].
- *L'évaluation informe, juge, aide à réorienter* [19].
- *Essentiellement un regard de rétroaction* [35].
- *Evaluer une action, c'est en produire le sens* [30].
- *Produire des connaissances sur lesquelles on applique un jugement pour prendre des décisions* [33].
- *Identifier ce qui est satisfaisant pour le maintenir et ce qui ne l'est pas pour l'améliorer* [2].

D'autres définitions plus précises sont indiquées dans le lexique en annexe. Quoiqu'il en soit, **trois éléments** se retrouvent dans toutes les définitions de l'évaluation de programme (Cf. schéma 3) : pour répondre aux questions d'évaluation, il faut une collecte d'informations, sur lesquelles on porte une appréciation critique, et qui débouche sur des propositions en vue d'une prise de décision.

Schéma 3 : Comment répondre aux questions d'évaluation ?



1. Une **collecte d'informations** de nature quantitative (mesures) et qualitative, avec un effort de collecte systématique et une recherche d'objectivité, ou en tout cas de diversification des points de vue. L'évaluation a donc d'abord une **dimension cognitive** [14],[12].
2. Une **appréciation critique** (un "jugement" de valeur), basée sur la comparaison de ces informations, sous forme de critères et d'indicateurs à des références ou "normes" (Cf. encadré 3). La deuxième dimension de l'évaluation est donc une **dimension normative** [14],[12]. Cette appréciation critique peut également être portée *en mettant en relation deux ou plusieurs objets du programme* [17]. Par exemple, en s'interrogeant sur la pertinence d'une action d'information pour atteindre un objectif de changement de comportement ; ou en s'interrogeant sur l'adéquation des ressources utilisées et des services produits.
3. Des **propositions** qui contribuent à une prise de décision. Ceci implique d'organiser l'évaluation (les méthodes, les objets à évaluer, les informations à recueillir) en fonction des questions auxquelles l'évaluation doit répondre, et ensuite de communiquer les résultats et les propositions aux décideurs, notamment sous forme d'un rapport, pour que la rétroaction se fasse. L'évaluation a donc une **dimension utilitaire** [12] ou *instrumentale* [14] : c'est un *instrument orienté vers la décision* [34]. Elle a aussi une *dimension démocratique* (elle alimente des débats) [12].

Toutefois, *les enseignements de l'évaluation ne sont pas les seuls éléments à jouer dans une prise de décision* [33] et parfois des modifications souhaitables ne dépendent pas des acteurs (décideurs, réalisateurs, concepteurs, partenaires) impliqués dans l'action. **La décision elle-même constitue un autre temps, et doit être séparée de l'évaluation.**

Encadré 3. Critères, indicateurs, normes.

- Un **critère**¹⁶ est un angle de vue porté sur un objet [19]. C'est une caractéristique observable [...] correspondant aux éléments du programme évalué [35]. C'est lui qui donne du sens à l'évaluation.
- Ces mêmes critères ne seront manipulables qu'une fois traduits en indicateurs. Un **indicateur** est un fait représentatif de la réalité [2], une variable qui aide à mesurer des changements [34]. C'est un instrument de mesure, qui donne de l'information. Il doit répondre à certaines qualités : il doit d'abord être valide (il mesure effectivement ce qu'il est censé mesurer [34]), mais aussi être repérable, tangible, éventuellement quantifiable, [...] disponible, [...] sensible et fiable (il doit varier dans les mêmes proportions que le phénomène qu'il mesure) [2].
- La **norme** (ou standard) est la valeur associée au critère [ou à l'indicateur] qu'on considère comme acceptable [9].

Exemples :

1. la mortalité est un critère de l'état de santé des nourrissons, le taux de mortalité infantile est un indicateur, et depuis les années 1990, la norme est de moins de 10 décès avant l'âge de 1 an pour mille naissances vivantes en France.
2. La pauvreté est un critère de la précarité, le revenu est un indicateur, la norme est moins de 60% du revenu médian de la population, soit 475 € de revenu par personne et par mois en France en 2003.

L'évaluation, pour faire quoi ?

Qu'elle soit formative ou sommative, l'évaluation est **un instrument de changement**. Son principal enjeu est de préparer des *changements, intentionnels et planifiés* : elle permet d'améliorer, d'adapter le programme [28]. Elle *donne à chaque acteur et partenaire les attributs nécessaires à une réelle évolution et/ou remise en cause* [8].

L'évaluation est d'une certaine manière une façon de questionner ce qui a été institutionnalisé dans une organisation. [Elle peut remettre] en cause un compromis établi dans le temps. L'évaluation, de par sa nature, sera

¹⁶ D'autres auteurs assimilent le critère à une norme en mettant en avant sa valeur numérique. Ce choix introduit à notre avis un risque de confusion.

toujours un moyen parmi d'autres de négocier les limites acceptables à l'intérieur desquelles le compromis reflété dans les pratiques pourra être questionné et remis en cause. En fait, c'est en produisant une déstabilisation que les connaissances produites par l'évaluation sont un ingrédient essentiel du changement dans les organisations [12] (Cf. encadré 4).

Encadré 4. Comment l'évaluation permet le changement.

L'évaluation favorise le changement par la prise en compte des recommandations et également par les débats qu'elle suscite entre les acteurs concernés. Il faut donc que l'évaluation ouvre à la fois une nouvelle compréhension de la réalité et une capacité à agir différemment :

- **Une nouvelle compréhension de la réalité** : *Le rôle de l'évaluation consiste précisément en la constitution d'une position "décalée" par rapport aux énoncés ordinaires qui [...] sont souvent trop massifs et volontairement flous (ou consensuels) pour donner lieux à débat contradictoire [14]. L'évaluation porte donc en soi un fort potentiel de controverses. Les acteurs ne peuvent ignorer totalement [les] informations [issues de celle-ci]. Même dans le cas où ils souhaitent les discréditer, ils doivent développer un argument leur permettant de les rejeter. Les acteurs doivent se doter d'un processus de délibération permettant de donner du sens à ces données et de s'entendre sur une interprétation viable pour la pratique. [...] La pratique de l'évaluation doit donc toujours être attentive à ce lien indissociable entre la production d'information et la production de la compréhension en milieu organisé [12].*
- **Une capacité à agir différemment** : *L'évaluation doit constituer un outil performant d'apprentissage et de mobilisation pour les acteurs des politiques publiques [14]. Par son effet de renforcement des acteurs, l'évaluation peut même devenir un élément de pérennité des équipes de prévention [36]. Quelle que soit la hauteur à laquelle on place l'enjeu, il est clair qu'on doit aussi évaluer pour soi, en tant qu'acteur, et pas seulement pour les autres : progresser en analysant les erreurs et les obstacles, capitaliser les expériences pour les réinvestir dans d'autres lieux ou d'autres programmes, est sans doute un enjeu motivant pour un professionnel si le contexte de l'évaluation le lui permet [18].*

Des dimensions stratégiques, techniques et sociales

L'acte d'évaluation ainsi défini appelle plusieurs remarques :

- L'évaluation est à la fois un ensemble de décisions ou de choix stratégiques, et une démarche méthodologique. Autrement dit elle est **enjeu politique et démarche technique**. Tout ce qui touche le cœur de l'évaluation est enjeu politique : choix de l'objet d'évaluation, des objectifs, du moment, des critères et indicateurs en fonction desquels sera portée l'appréciation, choix des acteurs de l'évaluation, des instruments de recueil de l'information, et utilisation des résultats. La dimension technique renvoie à la démarche et aux procédures d'évaluation, à la façon d'appliquer les techniques d'enquête, aux règles de construction des outils [...], aux méthodes de traitement des données et d'analyse des résultats [30].
- M. Demarteau insiste sur un troisième aspect de l'évaluation : l'évaluation est loin [...] d'être seulement une démarche scientifique, c'est aussi **une pratique sociale**. Ce qui importe est autant la rationalité accrue qu'elle apporte que la parole qu'elle libère [16].

Différencier l'évaluation de ce qu'elle n'est pas

- L'évaluation consiste en **une démarche rigoureuse et systématique** [17]. Cet effort d'explicitation différencie l'évaluation de l'analyse réalisée de manière intuitive et informelle par les intervenants dans le cadre de leurs interventions (Cf. encadré 5).
- Il n'y a pas d'évaluation sans appréciation critique. **Le suivi, la surveillance** (ou monitoring, monitoring), qui consiste à relever ou mesurer régulièrement des données (données de santé d'une population, ou déroulement d'un programme) sans les comparer à d'autres, ne sont pas de l'évaluation, parce qu'il manque cette comparaison à une norme qui permet l'appréciation [1],[17],[19],[35]. Cependant, la même information peut servir à la fois à la surveillance et à l'évaluation [35], et sans un bon suivi, il est difficile de faire une bonne évaluation [19]. Le suivi est en général basé sur un **tableau de bord**, qui regroupe l'ensemble des indicateurs recueillis. S'agissant du suivi d'un programme, les données sont en général regroupées dans un rapport d'activité.
- Il n'y a pas d'évaluation sans jugement, mais il s'agit d'une **appréciation critique** de situations données qui conduisent à formuler des conclusions judicieuses et des propositions utiles en vue d'une action ultérieure. Il ne s'agit pas de prononcer un jugement à la manière d'un

tribunal qui condamne les personnes pour des fautes [34]. Nous reviendrons sur cette différence essentielle entre contrôle et évaluation au chapitre suivant.

Encadré 5. L'évaluation : une pratique courante mais souvent intuitive.

Toute action planifiée implique au départ une certaine évaluation : engagé dans l'action, on ne peut pas ne pas évaluer ; on fait toujours le point sur l'action en cours. Dans ce sens, l'évaluation fait depuis longtemps partie des pratiques courantes. Mais cette pratique d'évaluation [...] reste souvent informelle. Bien que valable [...], elle apparaît limitée et inachevée [30].

Cela signifie que :

- Aucun organisme ne part de rien pour construire une évaluation des actions qu'il mène. L'enjeu est donc d'aider les organismes réalisant des programmes à *systématiser les pratiques évaluatives qu'ils ont développées dans l'action au fil du temps* [30].
- Malgré les efforts pour que l'évaluation soit aussi explicite et formelle que possible, il est rare que l'ensemble du processus (incluant l'utilisation qui sera faite des recommandations) le soit totalement. Il est donc fréquent qu'à un moment ou l'autre de l'évaluation ces aspects implicites et informels ressortent. Ils doivent alors être pris en compte.

Chapitre 3 : Des enjeux à prendre en compte dans l'organisation de l'évaluation

L'essentiel :

- L'évaluation est une démarche technique qui a des enjeux multiples et complexes, car elle s'inscrit dans un contexte politique, social et institutionnel, et elle interroge les logiques stratégiques des acteurs. La crainte de ces derniers peut être celle d'une évaluation qui n'en serait pas une, et poursuivrait un but de contrôle, voire de manipulation.
- Evaluer n'est jamais un exercice banal. La tension entre le besoin de reconnaissance et le risque de la critique est constitutive de l'acte d'évaluation. Une "véritable" évaluation est possible à condition d'être suffisamment transparente pour établir la confiance, désamorcer les craintes de contrôle et prendre en compte les attentes différentes qu'elle suscite.
- Cela suppose de disposer d'un espace d'autonomie par rapport aux décideurs de l'action, de construire et négocier l'évaluation avec les acteurs, de l'appuyer sur une répartition des rôles et une méthodologie explicites.
- Une démarche d'évaluation transparente, construite avec et négociée entre les principaux acteurs, est le meilleur moyen pour susciter la confiance et entraîner leur adhésion. Qu'elle soit interne ou externe, l'évaluation s'appuie toujours sur un travail collectif nécessitant la participation des acteurs du programme. Une répartition claire des rôles est un élément de la transparence de l'évaluation.

L'évaluation, telle qu'elle a été définie dans le chapitre précédent, est un exercice simple : il s'agit de se poser des questions et essayer d'y répondre. Néanmoins, elle est l'objet d'enjeux souvent importants pour les acteurs et les institutions concernées. C'est ce qui fait sa complexité.

Des enjeux multiples, explicites et cachés

Les questions posées par rapport au programme aboutissent à la commande d'une évaluation. Ces questions constitueront les objectifs de l'évaluation. Cependant, *l'évaluation, parce qu'elle est porteuse de transformations, peut faire l'objet de manipulation en fonction d'intérêts particuliers* [12]. C'est pourquoi elle déclenche souvent de la **méfiance** parmi les acteurs du programme. Ils craignent que l'évaluation soit en fait un prétexte pour exercer un **contrôle** sur eux, voire des **manipulations**.

- L'amalgame que font souvent les acteurs entre contrôle et évaluation tient essentiellement à la **crainte du jugement**. En effet, le jugement renvoie à une faute et à la condamnation d'une personne, et notre perception intuitive de l'évaluation est celle des notes reçues à l'école.

Evaluer le client¹⁷ signifie aussi, pour le travailleur social, évaluer sa propre pratique. Et là est l'ambiguïté de la démarche. Quand le travailleur social rend compte à son employeur (ou à un collègue, ou à tout autre service) de l'évolution de son client, il se met en position de vulnérabilité par rapport à eux. Il place autrui en position de dire : est-il un bon professionnel ou pas ? Entrent en jeu des mécanismes implicites (ou conscients) où le travailleur social met l'autre en position d'être son évaluateur. Ce qui l'oblige à donner une image positive de lui-même. Il peut être amené à mettre en œuvre des jeux stratégiques, quitte à tronquer la réalité, à faire passer au premier plan tel ou tel aspect et à relativiser tel autre. Personne ne souhaite donner de lui-même l'image d'un professionnel incompétent [10].

Rappelons que le "jugement" dont il est question dans l'évaluation est une appréciation critique portant sur des situations, et non sur des personnes (Cf. page 16). Il est donc essentiel de **différencier l'évaluation du contrôle** (Cf. encadré 6).

¹⁷ Le terme de "client" utilisé par l'auteur est ici équivalent à "usager".

Encadré 6. Ne pas confondre évaluation et contrôle.

L'évaluation et le contrôle s'opposent sur plusieurs points selon Y. Harvois :

- **L'évaluation** est une démarche **systémique et démocratique**, comportant de multiples niveaux de compréhension, pouvant questionner les vérités préétablies, visant à produire de la compréhension, centré sur l'expression de toutes les parties en présence avec leurs diversité. C'est un processus **accompagnant** l'action et débouchant sur "l'autre chose possible", le devenir, donnant l'autorisation politique à s'emparer de l'avenir, à évoluer, à aller vers **l'autonomie** [25].
- A l'opposé, **le contrôle** est une démarche **analytique et hiérarchique**, jugeant sur la base d'une norme préétablie ne pouvant être remise en question, où il y a séparation entre le contrôleur et le contrôlé, le premier ayant une supériorité dans la relation. C'est un processus conservateur, **sanctionnant** l'action et visant au maintien des acquis du présent et la **dépendance** à l'autorité¹⁸ [25].

De façon voisine, J. Beaudry pointe le risque de *dérive de l'évaluation vers le **contrôle bureaucratique***, qui consiste à s'intéresser à *des questions de performance et de rendement du personnel* ou à se mettre *au service de professionnels intransigeants* à qui elle permet de *s'assurer que les intervenants ne dérogent pas du plan d'intervention* [4].

- Plus grave encore, l'exercice appelé "évaluation" peut poursuivre des objectifs *implicites, inavoués ou même cachés*. Il s'agit alors de **manipulation** ou **pseudo-évaluation**¹⁹, telle que : *jeter de la poudre aux yeux* (autrement dit, faire de la publicité) ; *paralyser l'évaluation ou tenter de retarder l'évaluation véritable à plus tard [...]* en se livrant à un *exercice symbolique ou en déplaçant l'attention des vrais problèmes* ; *cacher certains aspects répréhensibles par des informations d'apparence favorables, véridiques ou mensongères* ; *exclure certaines personnes sous le faux prétexte d'une prétendue mauvaise performance* [3].

¹⁸ Cette présentation vise à bien différencier le contrôle de l'évaluation. Elle ne signifie pas que le contrôle est à rejeter. Il a un rôle régulateur important dans tout système car il permet l'application de la loi.

¹⁹ *L'évaluateur doit se garder de ces pseudo-évaluations, car il en payera toujours la note, soit que l'évaluation se fasse saboter (et il sera accusé d'incompétence), soit qu'il y perde lui-même son emploi (et il servira tout simplement de bouc émissaire)* [3].

D'où viennent les réticences, voire résistances à l'évaluation ? Au delà de la peur du contrôle, il y a souvent des **divergences** de perception sur ce qu'est l'évaluation et sur l'utilisation qui peut en être faite, selon les catégories d'acteurs concernées par le programme.

- Les **préoccupations** des acteurs peuvent être différentes. En général, les financeurs et décideurs sont d'abord intéressés par la *bonne utilisation des crédits* et veulent *constater une amélioration de l'état de santé d'une population* ou d'une situation sociale. Les acteurs des programmes veulent *être réellement utiles, mieux s'organiser pour être plus efficaces, faire savoir* [ce qu'ils font], *convaincre* de l'utilité de leur travail. La population cherche à *trouver une solution à ses problèmes* [33].
- Au delà des buts officiellement assignés à l'évaluation, les **buts officiels** des différents acteurs, souvent implicites, sont aussi très importants à considérer. Les administrateurs qui demandent une évaluation peuvent vouloir retarder une décision, légitimer une décision déjà prise, accroître leur pouvoir et le contrôle qu'ils exercent sur l'intervention, satisfaire aux exigences des organismes de tutelle. Les évaluateurs peuvent chercher à faire avancer les connaissances, mais aussi à accroître leur prestige et leur pouvoir, à obtenir une promotion, à promouvoir une idée qui leur est chère. Les utilisateurs peuvent chercher à bénéficier de services différents de ceux disponibles habituellement, à réduire leur dépendance vis-à-vis des professionnels. Le personnel d'une organisation peut chercher à court-circuiter les règles hiérarchiques, à obtenir un avancement [11].

L'évaluation se développe à partir d'un *postulat*, celui de la rationalité des activités que l'on analyse, car les acteurs sont soumis à des objectifs formels assignés par le programme [12]. Mais un programme est bien plus **complexe** qu'une série de ressources mobilisées pour mettre en œuvre des actions en fonction d'objectifs précis. Il est avant tout une **interaction** entre des "acteurs", personnes et organismes, qui sont impliqués pour son financement, son organisation, sa mise en œuvre, ou qui en sont bénéficiaires. Ces interactions sont complexes, mettant en jeu des logiques d'acteurs et d'institutions comme celles de pouvoir, de contrôle, de profit, voire de survie... Autrement dit, *la lecture que fait un acteur d'une situation est fonction d'une logique stratégique et détermine largement ses comportements*. L'évaluation fait donc face souvent au problème de devoir se positionner entre des logiques explicites et officielles et des logiques plus clandestines mais tout aussi marquantes [12].

Par exemple, un programme peut avoir pour objectif de développer l'estime de soi chez les enfants des écoles primaires, afin de prévenir l'échec scolaire et les conduites à risque à l'adolescence. Pour cela, des partenaires internes à

l'école (enseignants, service de santé scolaire) et externes (organismes de prévention) vont mettre en place des animations avec les enfants. Cette logique dictée par les objectifs va interroger d'autres logiques du système scolaire, telles que la place de l'école et le rôle des enseignants dans la santé mentale et l'éducation des enfants (notamment en rapport avec la place des parents), le rôle et la légitimité des intervenants extérieurs à l'école (notamment pour ceux qui ont une formation "psy"), ou encore la place des innovations à l'école.

En résumé, on peut dire que le programme (ainsi que son évaluation) s'inscrit dans un **environnement**, un contexte, dans lequel J. Beaudry différencie trois systèmes :

1. Le *macrosystème* est le **contexte politique**. Tout programme s'inscrit dans une action politique, parfois argumentée, contestable (visibilité politique, rassurer la population...). *Quelle que soit la raison d'être du programme à l'étude, toute évaluation comporte une dimension politique plus ou moins importante*. L'évaluateur devrait donc examiner s'il y a congruence entre ses actions et les objectifs qu'il poursuit en tant qu'agent de changement social avant de s'engager dans une évaluation [4].
2. L'*exosystème* est le **contexte social**. *L'évaluateur n'est qu'un des acteurs [...] et il n'y tient pas nécessairement le rôle principal. Politiciens, fonctionnaires, gestionnaires, praticiens, groupes de pression et usagers peuvent, à un moment ou à un autre, avoir droit de regard sur un projet d'évaluation. Ces derniers, ayant des intérêts particuliers à défendre et des besoins spécifiques à combler, pourraient être tentés d'influencer à leur avantage le cours de la démarche évaluative. L'évaluateur ne peut pas se permettre d'ignorer ces pressions. D'autre part, les normes et valeurs de la communauté dans laquelle s'effectue [l'évaluation] peuvent, à leur tour, exercer une influence directe sur la sélection des critères de réussite de l'intervention* [4].
3. Le *microsystème* est **l'organisme** de services [ou l'organisme qui conduit le programme]. *L'évaluateur [...] doit pouvoir compter sur la collaboration des intervenants activement engagés dans le programme. Sans leur participation, les chances de réussite de l'évaluation sont quasi-inexistantes. L'évaluateur ne peut donc pas ignorer ou mettre en veilleuse les appréhensions qu'ils manifestent à l'égard de son étude. Il peut s'agir : a) du refus des intervenants à transférer à la recherche des ressources normalement allouées à l'intervention; b) d'une crainte que l'évaluation ne remette en cause leur compétence et même leur poste; c) de l'expression du pessimisme des praticiens vis-à-vis de l'utilité ou la pertinence de l'évaluation; d) de leur inconfort face à l'évaluateur qui s'immisce dans leur travail et e) de leur refus de modifier les rôles et*

fonctions pour lesquels ils ont été engagés. La condescendance ou la négation du bien fondé de ces craintes risquent de ne pas convaincre. L'évaluateur a donc intérêt à considérer sérieusement les appréhensions des intervenants à l'égard de l'évaluation, à en discuter avec eux et à chercher des solutions propres à les satisfaire autant que lui-même [4].

Une tension saine et sage

L'évaluation doit composer avec cette complexité, dans laquelle elle se trouve elle-même incluse. En fait, c'est là que se situe son enjeu principal. *L'évaluation est d'autant plus productive qu'elle permet à des acteurs en milieu organisé de repenser les liens entre des objectifs qu'ils se donnent, des pratiques qu'ils développent et leur contexte d'action [12].*

Encadré 7. D'abord ne pas nuire

Le contrôle et les pseudo-évaluations sont des détournements évidents de la démarche d'évaluation.

Toutefois, une évaluation, même si elle est bien cadrée et organisée, peut avoir des effets délétères. Par exemple, un questionnaire d'évaluation de la qualité de vie peut être vécu comme intrusif dans l'intimité de la personne. Ou encore, s'intéresser à un groupe de personnes repérées comme ayant des besoins prioritaires peut renforcer l'exclusion sociale dont elles souffrent déjà.

Le premier principe éthique de toute intervention s'applique aussi à l'évaluation : d'abord ne pas nuire.

Pour L. Gaudreau et N. Lacelle, l'évaluation engendre toujours une tension entre :

- D'un côté **le danger** d'un détournement de mission, la quantité de travail à fournir, les enjeux de la connaissance et de pouvoir (jugements du bailleur de fonds, jugements sur les personnes...).
- De l'autre côté **les apports potentiels** : reconnaissance de notre travail, assurance que l'argent a été dépensé à bon escient, vérification de nos intuitions, découverte d'aspects inaperçus de notre travail, amélioration de nos interventions... [21].

*Cette tension de départ est saine et sage [et] ne doit jamais nous quitter parce qu'elle nous indique deux choses parfaitement essentielles. Elle nous dit [d'une part] de quel côté nous sommes, quel est le vrai "but du jeu" d'une évaluation : améliorer le sort de celles et ceux avec qui, pour qui, nous travaillons [...]. Elle nous dit [d'autre part] sur quel terrain de jeu nous mettons les pieds, où sont les buts, les filets à chaque bout de terrain pour marquer des points. L'évaluation nous apparaît instinctivement comme **un instrument de pouvoir et de connaissance**, et c'est ce qu'elle est. [...] Tout aussi instinctivement, on sait que cet instrument qu'est l'évaluation doit servir à la mission du groupe, "être profitable au monde". [21].*

S'organiser pour susciter la confiance

L'évaluation semble donc être une sorte d'utopie. Elle est basée sur la transparence, l'explicitation, la "démocratie", la recherche partagée d'une meilleure connaissance pour une meilleure performance. Alors que le système qu'elle étudie (et dont elle fait elle-même partie) est traversé sinon piloté par des enjeux contradictoires puissants, pour une bonne part implicites voire cachés, et par des conflits. L'exercice paraît difficile, et pourtant l'utopie peut fonctionner, sans se réduire à des généralités superficielles et inutiles. C'est **la confiance** qui le permet.

Cette confiance n'est pas acquise d'avance. Elle doit se gagner, d'abord par la création d'une **"zone d'autonomie"** de l'évaluation entre la décision, le contrôle et la connaissance pure [14]. Il s'agit d'élaborer un travail tenant compte à la fois des données objectives recueillies par l'évaluateur, croisées avec la demande du commanditaire, qui permette une prise de décision finale (poursuite de l'action, adaptation d'une nouvelle méthodologie, etc.). Cette zone d'autonomie doit contribuer à produire du débat et des échanges à partir d'un *processus de production de connaissances qui tiennent à l'égard de la politique une distance critique* [14]. L'élucidation des présupposés, la mise en discussion des catégories de l'action, de ses orientations et de ses finalités aident à la distance critique.

La confiance se gagne aussi par **l'explicitation** de l'ensemble de la démarche d'évaluation, c'est-à-dire :

- des objectifs de l'évaluation [3] clairement énoncés et partagés par les acteurs ;
- des critères de jugement, qui permettent d'afficher les valeurs qui sous-tendent ce qu'on considère comme "bien" ou "mal" [6] ;
- de la méthodologie et du processus d'évaluation (le déroulement de l'évaluation sur le terrain).

D'autres éléments seront développés sur les facteurs de qualité de l'évaluation (Cf. page 81). Mais d'ores et déjà, il est clair qu'une évaluation

ne doit pas être une parole d'expert, mais *un travail partenarial dynamique* qui permette de renforcer l'action [14]. **L'évaluation reste une affaire d'équipe.** Une démarche individuelle d'évaluation, concentrée dans le chef d'une seule et même personne, risque, à terme, d'être confrontée à un problème de légitimité [2]. La démarche d'évaluation inspire d'autant plus confiance aux acteurs du programme qu'elle est *participative et négociée* [21] :

- L'évaluation **participative**²⁰ permet de donner la parole aux membres de l'organisme et à ceux de la communauté dans laquelle il s'insère [21]. L'approche participative *accorde une place centrale à la participation de toutes les personnes concernées. Elle prend en considération leurs valeurs, leurs points de vue, leurs intérêts et leurs attentes, à toutes les étapes du processus d'évaluation* [30].
- L'évaluation est **négociée à l'interne** (la décision d'évaluer est prise après débats dans l'organisme), *et à l'externe* (pour que l'organisme évalué ait sa place en tant que partenaire de l'évaluation) [21]. *Parce qu'elle représente un enjeu politique, l'évaluation nécessite une négociation. La négociation consiste en un processus dynamique au cours duquel deux [ou plusieurs] parties ayant à la fois des intérêts communs et des intérêts opposés essaient de convenir des termes d'une entente* [30].

L'évaluation participative et négociée n'est pas seulement un prétexte pour déjouer les résistances à l'évaluation : l'implication des acteurs et des populations dans l'évaluation est en soi un moyen fort pour les rendre acteurs des changements qui sont dans les objectifs du programme.

D'ailleurs, l'opposition entre "chercheur" (ou évaluateur) extérieur à l'action et "acteur" impliqué dans l'action de routine devient parfois ténue, car on se situe souvent dans une démarche de recherche-action, d'expérimentation sociale, et alors il se crée *un pacte de connivence affective dans lequel le chercheur ne peut s'empêcher de devenir acteur et l'acteur s'aperçoit tout d'un coup qu'il est en train de chercher* [27].

²⁰ Ou *évaluation participante* [37], *évaluation pluraliste* [31].

Encadré 8. Faut-il toujours évaluer ?

Dans le principe, la réponse est oui. *Une évaluation est toujours nécessaire, même si son intérêt est parfois discuté [...]. L'évaluation est d'autant plus utile que : le thème est innovant dans le contexte local ; le programme est ambitieux ; l'acceptabilité n'est pas spontanée ; les résultats ou effets potentiels de l'action sur la population sont importants* [1].

Cependant, dans certaines circonstances, il est préférable de ne pas évaluer car *l'évaluation risque de faire plus de tort que de bien* [35] :

- Si l'évaluation pose des problèmes techniques importants ou a un coût élevé, alors que le programme semble bien aller et que la population y adhère [35].
- S'il n'y a pas de place pour des changements. Par exemple, lorsqu'on sait une activité peu efficace mais que le public souhaite la maintenir à tout prix ; ou lorsque les acteurs ne sont pas prêts à remettre en question leurs pratiques²¹ [35].
- S'il y a un conflit ou un problème entre les personnes ou les organismes impliqués dans le programme. Une démarche de *résolution de problème ou de gestion de conflit* est plus indiquée dans ce cas [21].

Différer l'évaluation ne signifie pas rester inactif, mais au contraire préparer activement les conditions de sa mise en place, par des actions de sensibilisation, de résolution de problèmes, de gestion des ressources humaines, etc.

Attention aux objectifs parasites : ni audit, ni publicité

D'autres objectifs que les buts formatif et sommatif sont parfois assignés à l'évaluation. Il faut les prendre en compte avec certaines précautions :

- Certains demandent à l'évaluation une **fonction comptable** [3], autrement dit le contrôle de l'utilisation des finances allouées [18],[31]. La bonne utilisation des fonds du programme est une question souvent posée à l'évaluation. Sa légitimité ne se discute pas, et les ressources sont un des objets de l'évaluation. Mais pour éviter toute ambiguïté entre contrôle et évaluation, il est préférable de bien séparer ce qui relève

²¹ Un travail préalable, soit avec le public, soit avec les acteurs, est alors nécessaire avant qu'une véritable démarche d'évaluation soit menée.

d'un **audit financier**²² et ce qui relève de l'évaluation du programme. Dans le même sens, nous déconseillons de parler de "**bilan**" pour une évaluation : un bilan est une analyse gestionnaire qui vise à établir la valeur monétaire d'une organisation et qui renvoie au concept d'audit.

- Des objectifs de **communication** sont parfois assignés à l'évaluation. Il s'agit par exemple de rendre visible le travail des associations ou groupes communautaires, de permettre à d'autres groupes, oeuvrant dans le même domaine, de justifier du même type d'action [37], ou encore de lever la suspicion sur l'efficacité et la légitimité d'un programme de promotion de la santé [18]. L'évaluation étant un exercice basé sur la diversification et l'explicitation des points de vue, et la restitution orale ou écrite d'informations, elle participe de fait à la communication. Néanmoins, cet objectif ne peut être central, au risque de dériver vers une sorte de publicité. C'est toujours la réponse aux questions soulevées par le programme qui constitue le cœur d'une évaluation.

Qui fait quoi ?

Compte-tenu de ce qui vient d'être dit sur les risques perçus d'une évaluation, le besoin de confiance dans son processus et la rigueur méthodologique nécessaire, la question du "Qui fait quoi ?" n'est pas anodine. Le choix de l'évaluateur ainsi que la répartition des rôles d'évaluation conditionnent le déroulement de l'évaluation et son succès. Deux types d'évaluation peuvent être différenciés :

- Dans une **évaluation externe**, *l'évaluateur est choisi en dehors de la structure responsable du projet* [31]. *L'extériorité de l'évaluateur est son principal avantage. Cette distance est perçue en effet comme un gage d'objectivité, de neutralité, voire de liberté par rapport aux pressions éventuelles. Ses compétences d'expert sont garantes de la rigueur scientifique du travail accompli et il peut jouer un rôle de médiateur entre les divers responsables et acteurs du programme* [2]. L'évaluation externe est donc considérée comme la plus valide.

Toutefois, elle *court le risque d'être déconnectée des préoccupations des acteurs* [19], car l'évaluateur externe connaît mal le programme et peut être *tenté par une démarche trop théorique au service de ses propres recherches*. De plus, l'extériorité de l'évaluateur *ne garantit pas*,

²² L'audit est un contrôle de légalité et de conformité par rapport aux normes de fonctionnement définies par les lois et les règlements, directives, budgets, circulaires, notes internes [31].

de sa part, l'absence de tout préjugé ou idée préconçue sur la santé ou la communauté concernée, et il a un coût qui doit être assumé[2].

- Dans une **évaluation interne**²³, l'évaluateur appartient à la structure responsable du projet [31]. Elle est réalisée par les responsables [...] de l'action [19]. Néanmoins, ils peuvent être assistés par un expert externe. L'évaluation interne a l'avantage de s'appuyer sur ceux qui connaissent le mieux le programme et la population visée. Les évaluateurs perçoivent mieux les problèmes rencontrés et ils possèdent, en outre, une bonne connaissance du programme en cours. Cette proximité permet dès lors d'entretenir plus aisément une dynamique d'évaluation permanente destinée à améliorer la dynamique du projet, le processus en cours [2].

Toutefois, trop de proximité [...] peut conduire l'évaluateur à ne plus prendre suffisamment de distance par rapport à l'objet de son travail [2], et l'évaluation peut tourner à l'autosatisfaction [19]. Enfin, le temps et les moyens nécessaires à l'évaluation peuvent manquer à l'évaluateur, ainsi que certaines connaissances méthodologiques de base [2] (Cf. encadré 9).

L'autoévaluation est une modalité particulière d'évaluation interne²⁴, pratiquée par des "bénéficiaires" de l'action ou leurs représentants. Elle constitue pour ce groupe d'acteurs une méthode d'appropriation de l'action, et ils l'évaluent selon leurs propres critères [19]. Il ne s'agit pas de l'évaluation que formulent les bénéficiaires de l'action d'un autre, mais bien de l'évaluation d'un groupe communautaire et de son action par lui-même [2]. Il peut cependant être assisté par un "accompagnateur" externe.

²³ Ou *endo-évaluation*, évaluation par une personne "du dedans", par opposition à l'*allo-évaluation*, ou évaluation par une personne extérieure [6].

²⁴ Tous les auteurs ne distinguent pas cette modalité, beaucoup considèrent l'autoévaluation et l'évaluation interne comme synonymes.

Encadré 9. Participation et objectivité sont compatibles

Si les [...] acteurs participent à l'évaluation de leur propre travail, s'ils sont partie prenante du jugement sur la valeur de ce qu'ils font, n'y a-t-il pas de danger qu'ils soient à la fois juge et partie ? Comment garantir alors l'objectivité du jugement et se prémunir contre les biais de la subjectivité ? Comment imprimer un caractère scientifique à la démarche d'évaluation ? [30].

Dans une évaluation, le "jugement" de valeur n'est pas objectif, si l'on entend par là qu'il serait neutre, subjectivement désengagé. Il n'est pas non plus subjectif, dans le sens qu'il serait dénué de tout fondement. Au contraire, dans toute évaluation systématique, le jugement se base sur des informations qui sont méthodiquement recueillies avec toute l'objectivité requise. L'évaluation est donc à la fois marquée par l'objectivité et la subjectivité. L'évaluation est une démarche où est investie la subjectivité des [...] évaluateurs, et qui est régie par les règles d'objectivité de l'enquête sociale [30].

En conclusion, la responsabilité de l'évaluation doit être assumée de façon différente, c'est-à-dire par des personnes différentes, selon l'élément du programme que l'on veut évaluer et selon le but poursuivi par l'évaluation. Ainsi, pour une évaluation sommative, c'est-à-dire visant à déterminer les effets d'un programme, la place du chercheur, externe au processus d'implantation, est prépondérante. Par ailleurs, lors de l'implantation du programme, alors que la recherche formative est particulièrement appropriée, la contribution des personnes responsables de la mise en œuvre du programme est essentielle. [...]. Plusieurs auteurs sont donc d'accord pour conclure que l'évaluation devrait réunir idéalement, et les chercheurs responsables de l'évaluation, et les gestionnaires directement responsables de la mise en œuvre du programme [35].

Ces questions d'organisation de l'évaluation et de méthode seront vues plus en détail dans la seconde partie (chapitres 4 et 5). Il faut auparavant préciser sur quels "objets" l'évaluation va porter.

Chapitre 4 : Evaluer quoi ? Les objets de l'évaluation

L'essentiel :

- Une fois définis les objectifs de l'évaluation et les questions posées à l'évaluation (chapitre précédent), il faut définir les objets du programme qui seront concernés par l'évaluation et le type d'évaluation qui sera faite.
- L'ensemble de ces éléments constitue le "cœur" de l'évaluation, c'est-à-dire la stratégie d'évaluation à mettre en œuvre au moyen d'une méthodologie.
- Un programme a cinq niveaux, qui résument la logique de l'intervention et qui peuvent être objets d'évaluation :
 1. Des ressources nécessaires à sa réalisation (évaluation de la structure).
 2. Des activités ou services à réaliser (évaluation du processus).
 3. Des groupes ciblés par les activités ou services (évaluation d'implantation).
 4. Des objectifs ou résultats attendus dans ces groupes-cibles (évaluation du résultat, efficacité ou utilité).
 5. Un impact, ou des retombées (évaluation d'impact).
- Les évaluations portant sur les trois premiers objets s'intéressent aux efforts des acteurs pour mettre en œuvre ou implanter le programme. Elles sont les plus importantes dans un but formatif.
- Les évaluations portant sur les deux derniers objets s'intéressent aux effets du programme dans les groupes-cibles ou bénéficiaires, et au delà. Elles sont les plus importantes dans un but sommatif. L'évaluation des effets ne prend du sens que si elle est reliée à la mise en œuvre du programme.

Dans ce chapitre sont abordés les objets de l'évaluation, c'est-à-dire les parties des programmes qui peuvent faire l'objet d'évaluation, et les différents types d'évaluation qui s'y rapportent.

L'évaluation peut porter soit sur l'ensemble d'un programme, soit sur une partie de celui-ci²⁵.

Un programme peut se décomposer en cinq niveaux pouvant chacun faire l'objet d'une évaluation.

Les cinq objets d'évaluation d'un programme

Il est classique, tout d'abord, de différencier trois objets d'évaluation principaux : les ressources, les activités ou services, les objectifs. A chaque niveau correspond un type d'évaluation :

1. **Ressources ou structure** (ou "input", entrées, intrants). **L'évaluation de la structure** s'intéresse à *l'adéquation des ressources ("adequacy") : la quantité, la qualité et l'agencement des ressources humaines, physiques et financières du programme* [35]. On parle aussi d'*évaluation des ressources* [29]. La structure renvoie donc aux ressources mais aussi à *l'organisation d'ensemble avec ses mandats, ses règles et son fonctionnement* [17].

Exemple : pour réaliser un programme d'échange de seringues en pharmacie (à destination des usagers de drogues par voie injectable), une association de réduction des risques a besoin des **ressources** suivantes :

- Disposer de personnel pour réaliser le programme en partenariat avec des officines et d'autres partenaires (financeurs, etc.).
- Disposer d'un véhicule pour transporter les kits d'injection et les conteneurs de déchets.
- Disposer d'un financement pour l'achat des kits, des conteneurs, et pour le fonctionnement du personnel associatif mobilisé pour ce programme.

²⁵ Par analogie avec la chimie, on peut différencier *l'approche molaire* si le programme est évalué dans sa globalité, et *l'approche moléculaire* si l'évaluation est limitée à *certaines stratégies ou actions spécifiques* [4]. On peut également parler de méthode des *traceurs* pour cette approche partielle [24], qui consiste en une sorte de sondage : l'évaluation se focalise sur une partie qui semble représentative de l'ensemble du programme.

2. **Activités ou services** (ou processus, "output", sorties, extrants). *L'évaluation du processus* s'intéresse à la justesse (*appropriateness*) des activités. Dans le cas d'un service, on s'intéresse à trois aspects : la qualité de ses *aspects techniques* [...] (*bonnes pratiques*), la qualité de la relation entre le bénéficiaire et le professionnel, et les *processus organisationnels* [35]. On parle aussi d'*évaluation des procédures* [5],[28], terme qui est plus adapté aux procédures de soins.

Exemple : les principales **activités** du programme d'échange de seringues en pharmacie sont le recrutement des pharmacies et la mise en place d'un protocole d'échange de kits, la sensibilisation et la formation des équipes officinales, l'approvisionnement en kits et conteneurs, l'échange proprement dit des kits neufs contre les kits usagés, puis la récupération et l'élimination des conteneurs de déchets.

3. **Objectifs**. *L'évaluation des résultats* se rapporte aux objectifs du programme. Les services produits ("output") ont-ils provoqué les changements attendus ("outcome") ? Par "résultats" on entend généralement des changements directement liés à la raison d'être du programme, c'est-à-dire des changements liés aux problèmes visés par le programme et pour lesquels celui-ci a été élaboré [9]. S'agissant de programmes de développement des services de santé, les objectifs sont définis en terme d'état de santé ou de comportements relatifs à la santé de la part des bénéficiaires visés ou atteints par le programme [35]. Mais tous les programmes n'ont pas directement des objectifs de santé : il peut s'agir d'objectifs de connaissance, de pratique professionnelle, d'utilisation d'un service...

Exemple : l'**objectif spécifique** du programme d'échange de seringues en pharmacie est de réduire la prise de risque infectieux (notamment VIH/sida et hépatites virales B et C) chez les usagers de drogues par voie injectable par l'utilisation systématique de matériel de préparation et d'injection stériles.

La seule approche quantitative des résultats n'est pas suffisante. Une approche plus analytique des résultats amène à se poser les questions plus précises telles que : Quels sont les aspects du programme qui sont efficaces ? Comment et en quoi sont-ils efficaces ? Pour qui sont-ils efficaces ? [6].

Encadré 10. Ne pas confondre efficacité et utilité.

La distinction entre bénéficiaires visés et bénéficiaires atteints par le programme permet de différencier l'efficacité et l'utilité. *Dans le cas où le dénominateur est l'ensemble de la population visée par le programme (population-cible), l'on parle d'évaluation de l'efficacité²⁶ ("effectiveness") du programme. S'il s'agit seulement de la population rejointe par le programme, l'on parlera d'évaluation de l'utilité²⁷ ("efficacy").* [35].

Exemple :

- L'utilité du programme d'échange de seringues en pharmacie est mesurée par le pourcentage des usagers de drogue injectable (UDI) qui utilisent systématiquement du matériel stérile par rapport à tous les UDI qui rentrent dans le protocole d'échange de seringues.
- L'efficacité du même programme est mesurée par le pourcentage des UDI qui utilisent systématiquement du matériel stérile par rapport à tous les UDI habitant dans la zone du programme.

La relation entre les deux dépend en effet d'une part du taux de pénétration dans la population visée (% de la population rejointe), et du taux d'observance de l'action de santé par cette population et les fournisseurs de soins (% d'observance), si bien qu'on peut poser l'équation suivante [35] :

$$\text{Efficacité réelle} = \text{utilité} * \% \text{ pénétration} * \% \text{ observance}^{28}$$

Deux autres objets d'évaluation peuvent compléter les trois qui viennent d'être décrits : les populations-cibles et l'impact :

4. **Populations-cibles.** Dans l'aspect organisationnel de l'évaluation du processus, F. Champagne inclut le fait de savoir *si le programme atteint les populations visées* [9]. A l'inverse, J. Beaudry préfère parler de ***l'évaluation d'implantation***²⁹, qui conduit [...] à l'identification de la

²⁶ L'efficacité dans des situations concrètes et réelles où des contraintes existent peut encore s'appeler *efficacité réelle* [35] ou *efficacité populationnelle* [11].

²⁷ L'utilité dans des conditions optimales peut encore s'appeler *efficacité potentielle* [35]. De façon plus précise, *l'efficacité d'utilisation* concerne la population-cible rejointe dans les conditions du terrain (donc avec une observance plus ou moins grande), alors que *l'efficacité théorique* et *l'efficacité d'essai* correspondent à des situations où l'intervention est contrôlée, la première en laboratoire et la seconde dans un essai expérimental [11].

²⁸ Le signe * indique une multiplication.

²⁹ L'implantation d'une action ou d'un service dans la population-cible est aussi appelée *couverture* [12],[28],[32] ou *pénétration* [35].

population qui fait effectivement usage des services mis en place [4]. Il souligne ainsi l'importance de rejoindre la population-cible : la participation de la population-cible constitue une mesure critique de la validité d'un programme. Toutefois, il englobe aussi dans l'évaluation d'implantation un second volet qui s'intéresse aux ressources et aux activités [4]. Par souci de clarté, nous différencions ici l'évaluation d'implantation prise dans son sens premier (la population atteinte) de l'évaluation du processus et de la structure (qui s'attachent respectivement aux activités et aux ressources).

Exemple : la **population-cible** du programme d'échange de seringues en pharmacie est celui des usagers de drogues injectables. Comme ils sont touchés par le biais des pharmacies qu'ils fréquentent, les équipes officinales sont une population-cible intermédiaire pour les activités de support et d'approvisionnement.

5. Conséquences, retombées, ou **impact** du programme : *Alors que les résultats sont des changements directement liés aux objectifs du programme, l'impact inclut un grand nombre de changements directs et indirects, voulus ou non voulus, liés ou non aux objectifs explicites du programme, tant sur la population visée que sur la population entière. L'évaluation de l'impact peut donc porter sur l'état de santé de la population, sur l'utilisation des services de santé en général ainsi que sur des phénomènes environnementaux, économiques, sociaux, etc. [9].*

Exemple : **l'impact** attendu du programme d'échange de seringues en pharmacie est à plusieurs niveaux :

- Au niveau sanitaire, une réduction des nouvelles infections (notamment VIH/sida et hépatites B et C) chez les UDI, mais peut-être aussi une augmentation du recours aux traitements de substitution.
- Au niveau social, une meilleure perception réciproque des pharmaciens et des UDI, et plus généralement un regard social moins stigmatisant pour les UDI.
- Au niveau environnemental, moins d'objets piquants-coupants dans les ordures ménagères et les lieux publics.

L'impact³⁰ est donc une approche large des effets du programme : il comprend à la fois la dimension temporelle du long terme³¹, la

³⁰ Dans les campagnes de communication, le terme "impact" est utilisé pour désigner le nombre de personnes qui se souviennent d'avoir vu le film, l'affiche, ou entendu le spot radio (Voir : Roussille B., Arwidson P. **L'éducation pour la santé est-elle efficace ?** Vanves : CFES, 1998, 45 p.). Ce terme est toutefois impropre puisqu'il s'agit de savoir si la population

dimension populationnelle, c'est-à-dire les effets sur une population plus large que la population rejointe ou la population-cible et, finalement, la dimension non spécifique des effets, c'est-à-dire des conséquences plus larges que les résultats spécifiques du programme, comme les conséquences du programme pour le développement sanitaire et social [35]. Ainsi, il est nécessaire de concevoir aussi des systèmes d'évaluation qui permettent de mesurer les effets non attendus (positifs ou pervers) d'un programme, en plus des effets prévus [15].

Par exemple, le retrait préventif des femmes enceintes de milieux de travail dangereux pour le fœtus peut donner des résultats en terme de réduction de la morbidité et mortalité périnatale, mais peut avoir pour impact une discrimination des femmes pour certains emplois.

Les questions de la pérennité des effets, de la viabilité des mécanismes d'action mis en place sont également une composante de l'évaluation d'impact. Ils permettent de répondre à la question de la répliquabilité du programme, à savoir la possibilité de reproduire ailleurs ce qui a été validé dans un contexte donné [19],[31]. *La viabilité s'apprécie en termes institutionnels, financiers et techniques [31].*

En résumé, un programme est constitué de **cinq éléments** qui sont, dans l'ordre de leur mise en place sur le terrain : des ressources ("input"), permettant de réaliser des activités ou services ("output") en direction d'un ou plusieurs groupes-cibles, qui vont donner des résultats (autrement dit, atteindre leurs objectifs) ("outcome") et d'autres conséquences (impact).

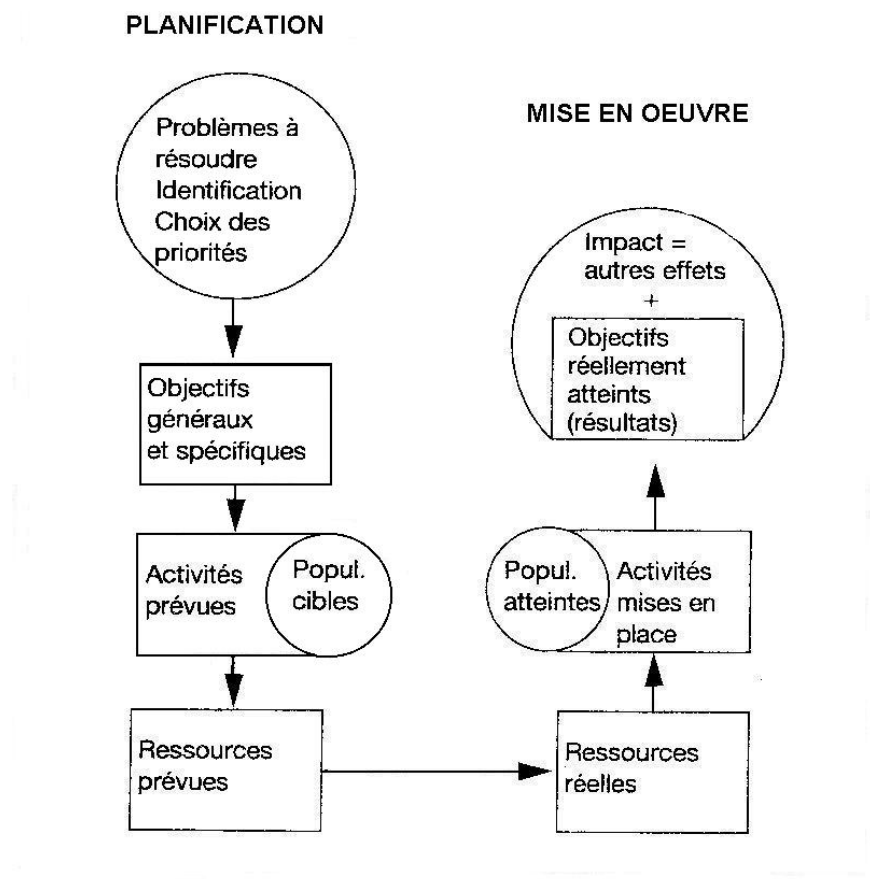
Ces éléments constituent le cadre logique du programme. Ils peuvent être figurés sur un schéma qui récapitule les phases de planification et de mise en œuvre du programme (Schéma 4).

ciblée a été rejointe : selon la terminologie adoptée ici, il s'agit de l'implantation (ou de la couverture).

³¹ La plupart des auteurs adoptent cette distinction entre résultats et impact, mais certains les entendent dans le sens opposé (impact = effet à court terme, résultat = effet à moyen ou long terme)...

Schéma 4 : Les "objets" du programme : étapes de la planification et de l'action.

Source : Fontaine D., 1992 [20]



Les cinq grands types d'évaluation, vus ci-dessus, ne suffisent pas à répondre à toutes les questions de l'évaluation. En effet, évaluer chaque "objet" du programme ne suffit pas : il faut aussi évaluer les liens entre ces "objets" ou composantes. De plus, certains types d'évaluation peuvent être regroupés.

A nouveau, il faut revenir aux questions posées. On peut regrouper les types d'évaluation selon les deux grands buts de l'évaluation (Cf. page 20) : l'évaluation formative (ou concurrente, concomitante, continue, permanente, dynamique, en cours) et l'évaluation sommative (ou récapitulative, globale, intermédiaire ou finale).

Evaluer les efforts

L'évaluation formative a un objectif de rétroaction rapide en vue du pilotage, de l'amélioration, de l'adaptation de la gestion du programme au cours de sa mise en œuvre. Cette évaluation est centrée sur *les efforts* déployés par les acteurs du programme pour mobiliser les ressources, réaliser les activités, et rejoindre les groupes-cibles concernés [20]. Les évaluations de la structure, du processus, et d'implantation sont donc les plus intéressantes dans ce cas (Cf. Schéma 5).

Ces trois évaluations peuvent être regroupées sous le terme d'**évaluation de la mise en œuvre** [17], ou d'*évaluation d'implantation* prise au sens large³² [4]. Cette évaluation *s'intéresse au déroulement réel du programme [et] permet de connaître la façon dont les ressources ont été utilisées, si les activités prévues ont été réalisées, les produits et les services offerts (extrants), la satisfaction [et] le degré d'atteinte [les publics visés], etc. De façon plus globale, elle s'intéresse à la dynamique interne du programme et aux facteurs tant internes qu'externes qui influencent sa mise en œuvre* [17]. Elle devrait faire partie des tâches routinières de tout organisme [4].

On peut différencier deux types d'évaluation de la mise en œuvre (ou d'implantation), selon la référence utilisée pour l'appréciation :

- Dans le cas où les ressources, activités et groupes-cibles sont évalués en référence à la programmation prévue, il s'agit de **l'évaluation de l'écart**³³ entre le *programme "idéal", c'est-à-dire le programme tel qu'on l'a planifié* et [le] *programme "actuel" c'est-à-dire tel qu'on l'a implanté*.

³² Ce sens large étant répandu, nous l'avons conservé. Nous préférons cependant l'employer dans son sens restreint, qui est l'atteinte des populations-cibles du programme (Cf. page 43).

³³ Ou évaluation du déroulement de l'action [1], évaluation opérationnelle [35], évaluation normative (qui inclut également l'atteinte des objectifs) [12].

[Elle] est essentielle car souvent les programmes réels sont significativement différents de ceux qui avaient été prévus. La réalité d'exécution et d'implantation des programmes provoque souvent des ajustements et changements majeurs [9].

Exemple : dans le programme d'échange de seringues en pharmacie, l'évaluation de l'écart s'intéresse à la mobilisation réelle des ressources (de l'association comme de celle des partenaires) et à la mise en place des activités suivant le calendrier retenu, en commençant par le recrutement des pharmacies volontaires, leur formation, la mise en place du protocole d'échange de seringues, puis ensuite l'activité d'échange de seringue réellement menée, par rapport à l'activité qui était attendue en raison de la file active estimée.

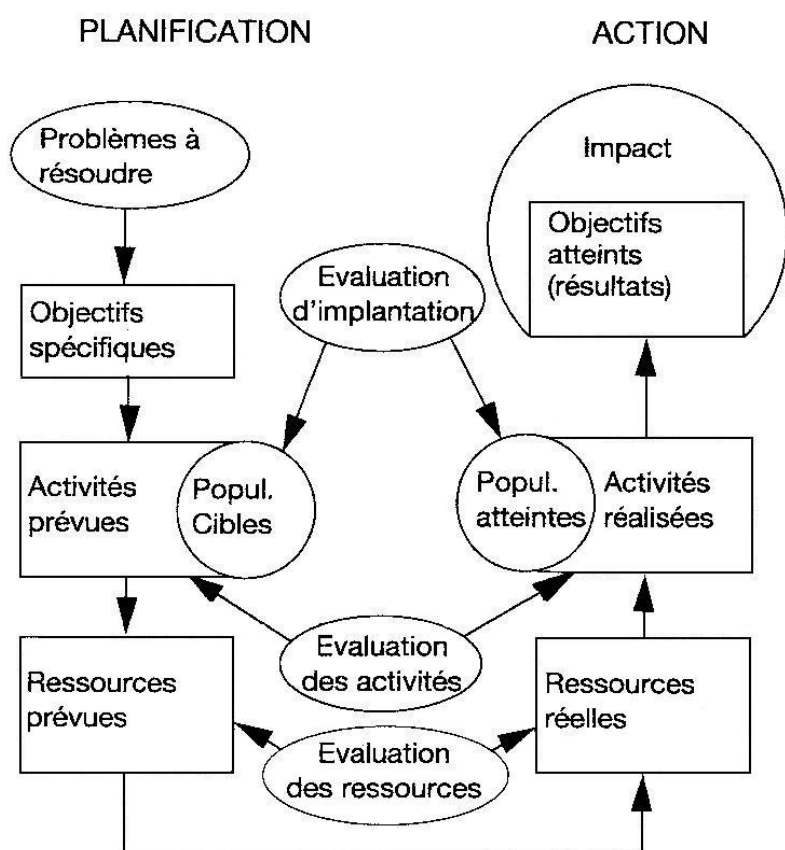
- Dans le cas où les ressources, activités et groupes-cibles sont évalués en référence aux objectifs poursuivis (donc à la stratégie d'action), il s'agit de ***l'évaluation de la cohérence***³⁴ [1],[31],[33]. *Les moyens, ressources, interventions mis en place vont-ils permettre de remplir les objectifs ?* [33]. Autrement dit, le programme a-t-il été implanté en cohérence avec ses objectifs ?

Exemple : dans le programme d'échange de seringues en pharmacie, l'évaluation de la cohérence s'interroge sur la capacité des ressources mobilisées dans le cadre du programme à mener toutes les activités nécessaires pour mobiliser, former et approvisionner les équipes officinales, pour que celles-ci puissent proposer l'échange de seringues aux usagers de drogues injectables de façon à retenir leur attention et qu'ils disposent du matériel stérile et de l'information leur permettant d'utiliser au mieux ce service d'échange de kits.

³⁴ Ou *analyse du programme* [9], *analyse logique de l'intervention* [12], ou encore *évaluation tactique* [35], qui inclut l'évaluation des objectifs spécifiques (évaluation des résultats).

Schéma 5 : les évaluations centrées sur les efforts ou la mise en oeuvre

Source : Fontaine D., 1992 [20]



Evaluer les effets

L'évaluation sommative s'intéresse aux effets du programme, c'est-à-dire *tout comportement ou tout événement dont on peut raisonnablement dire qu'il a été influencé par un aspect quelconque du programme* [7].

L'évaluation des résultats et l'évaluation d'impact, déjà vues précédemment (Cf. pages 42 et 44), sont donc les deux principales évaluations à faire dans un but sommatif. Elles permettent de répondre aux deux critères classiques que sont l'efficacité et l'impact :

- **L'efficacité.** Le programme a-t-il atteint ses objectifs ? Il s'agit de mettre en lien l'atteinte des objectifs (l'évaluation des résultats) d'une part avec les activités qui ont été implantées et d'autre part avec les objectifs prévus avant le lancement du programme. On a déjà différencié l'efficacité et l'utilité selon la population concernée (Cf. page 43).
- **L'impact.** L'ensemble des conséquences du programme, dont on a souligné la globalité, la durée dans le temps, et l'extension à plusieurs populations. Là encore, il s'agit de mettre en lien les activités et les conséquences.

La distinction entre résultats et impact n'est pas toujours très pertinente, aussi peut-on les regrouper en **évaluation des effets**³⁵, qui s'intéresse au *degré d'atteinte des résultats escomptés et des autres conséquences et porte sur les résultats spécifiques obtenus par le programme en regard des objectifs pré-établis ou encore sur les conséquences ou l'impact du programme* [35].

Deux autres critères sont classiques dans une évaluation sommative, l'efficacité et la pertinence :

- **L'efficacité**, ou rendement ("efficiency"), est une évaluation économique qui met en relation les résultats obtenus et les ressources mobilisées. Autrement dit elle est basée sur le calcul du coût unitaire du résultat. Dans ce document, nous n'avons pas traité ce type d'évaluation qui nécessite des analyses de coûts poussées et une homogénéité des résultats. La question qui est en fait posée à l'évaluation est plutôt celle-ci : les résultats obtenus l'ont-il été à un coût acceptable ? La question se rapproche de celle de la cohérence entre les ressources et le résultat. La réponse n'est donc pas un chiffre mais un ordre de grandeur.

³⁵ Certains auteurs utilisent le terme **d'évaluation d'impact** au sens large pour l'évaluation de tous les effets d'un programme, en incluant les résultats.

Exemple : dans le programme d'échange de seringues en pharmacie, l'évaluation de l'efficacité s'intéresserait au coût du kit neuf utilisé par un usager de drogue injectable avec destruction d'un kit usagé.

- **La pertinence.** Le programme existe en fonction d'une situation problématique identifiée et analysée lors de la phase de planification. *Quelle peut être la contribution du programme à la solution des problèmes identifiés ?* [35]. **L'évaluation stratégique** vise à apprécier l'adéquation entre le programme et les problèmes à résoudre. En d'autres termes, il s'agit de juger de la pertinence des objectifs du programme en regard des problèmes qui affectent la population en question [35].

Cette évaluation peut également s'appeler **évaluation de la pertinence**³⁶ et R. Pineault et C. Daveluy différencient l'évaluation de la pertinence théorique qui se situe *ex-ante*, au moment de la planification et l'évaluation de la pertinence réelle ou empirique qui se situe à la fin du programme [35]. La pertinence réelle prend en considération tous les effets du programme, résultats et conséquences. Elle s'alimente pour ainsi dire à l'évaluation des résultats et à l'évaluation d'impact, en les inscrivant dans un contexte de pertinence par rapport au contexte plus large dans lequel les problèmes de santé sont posés et les priorités établies. Il y a donc un lien très étroit entre l'évaluation d'impact et l'évaluation de la pertinence réelle [35]. Elle met donc en relation l'impact du programme (incluant les résultats) et la situation problématique qui l'avait justifié. Cette conception [de la pertinence] suppose d'avoir développé une capacité d'analyse et de diagnostic social [10] avant le début du programme.

Exemple : dans le programme d'échange de seringues en pharmacie, l'évaluation stratégique s'interroge sur la pertinence de la distribution des kits stériles en officine en retour de matériel usagé pour lutter contre l'épidémie d'infections à VIH/sida et d'hépatites virales B et C qui touche fréquemment les usagers de drogues injectables.

Les trois principales évaluations dans un but sommatif sont donc **l'évaluation des résultats**, **l'évaluation d'impact**, et **l'évaluation de la pertinence** (ou évaluation stratégique) (Schéma 6).

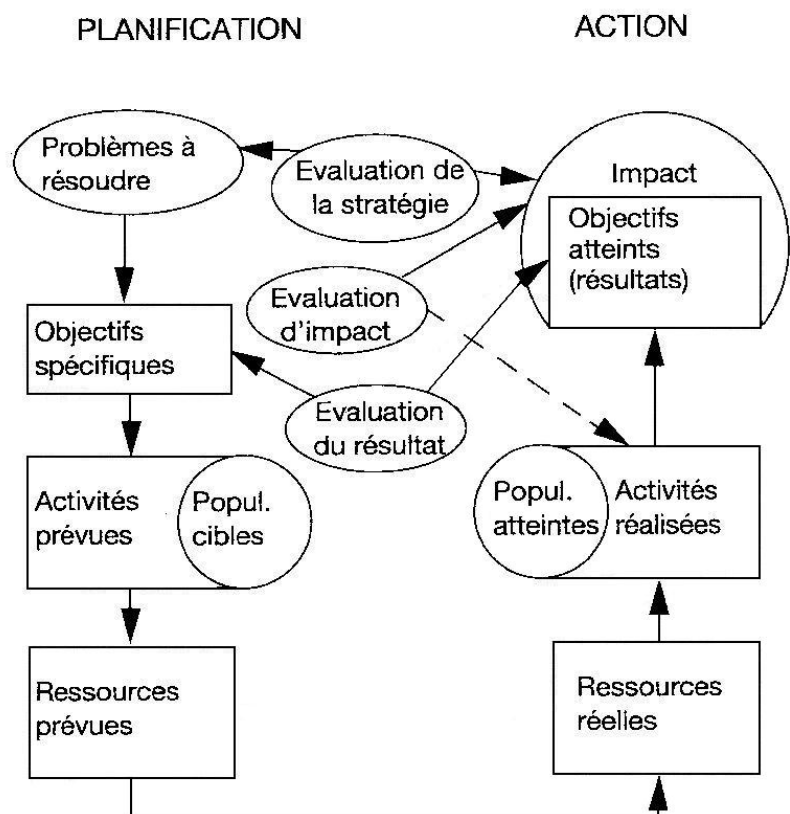
L'évaluation des effets est *celle qui retient le plus d'attention*. Cependant, il faut d'abord s'intéresser aux ressources et aux activités, car c'est en les reliant aux effets que l'évaluation prend son sens. *Si le programme étudié ne*

³⁶ R. Pineault et C. Daveluy mettent en garde contre l'utilisation du terme *évaluation de l'adéquation* pour l'évaluation de la pertinence, car l'adéquation devrait être réservée aux ressources, donc à l'évaluation de la structure [35].

rejoint pas la population cible pour laquelle il a été conçu ou si les opérations prévues ne sont pas appliquées de façon conséquente avec les objectifs de l'intervention, on ne peut pas s'attendre à ce que ses résultats [...] soient pertinents ou même sensés [4]. Autrement dit, l'évaluation des efforts est une étape préalable à l'évaluation des effets.

Schéma 6 : Les évaluations centrées sur les effets

Source : Fontaine D., 1992 [20]



2^{ème} PARTIE :

LA PRATIQUE DE L'EVALUATION :

METHODES ET ETAPES

Chapitre 5 : Comment évaluer ? La méthodologie

L'essentiel :

- Chaque programme est unique par ses acteurs, ses objectifs, ses activités et son contexte. Il n'y a pas de "recette" en évaluation.
- Une bonne méthodologie comporte à la fois des éléments d'analyse selon la rationalité des objectifs du programme (modèle axé vers les buts) et des éléments d'analyse selon les rationalités des acteurs dans leur environnement (modèle systémique).
- Le modèle axé vers les buts s'inscrit dans les méthodes de la recherche en épidémiologie. Il cherche à mettre en évidence un lien entre une intervention et des effets. Il est souvent plus approprié pour l'évaluation des résultats.
- Le modèle systémique s'inscrit dans les méthodes de la recherche en sociologie. Il est souvent plus approprié pour l'évaluation du processus et de l'impact.
- Le recueil d'information fait appel à la recherche documentaire, à l'observation, aux entretiens et aux enquêtes par questionnaires ou par entretiens.
- Les points clés de la méthode d'évaluation concernent les "mesures" (validité des indicateurs) et l'appréciation critique portée sur celles-ci (construction des critères et des normes en référence à un système de valeur). Ceci donne d'autant plus d'importance à la rigueur méthodologique.
- La diversification des sources de données et des méthodes de recueil, ou "triangulation", est un élément important pour renforcer la validité de l'évaluation.

Les méthodes d'évaluation dérivent des méthodes de la recherche : soit de la recherche épidémiologique, soit de la recherche en sciences sociales. Ces deux champs correspondent à deux modèles ou paradigmes différents du programme et de son évaluation. Le programme est-il considéré d'abord comme une entité finalisée et rationnelle, ou un système dynamique complexe ? Privilégie-t-on l'observation extérieure ou les interactions ? L'objectivité ou les subjectivités ? L'approche quantitative ou qualitative ? [35],[14]. On peut concevoir le programme soit comme une entité rationnelle axée vers les buts, soit comme un système complexe d'interactions (Cf. encadré 11).

Encadré 11. Deux modélisations du programme.

- *Dans le premier modèle, l'organisation est conçue comme une entité simple poursuivant un ou des buts facilement identifiables. Toutes les activités sont orientées vers la poursuite de ces buts et l'évaluation [...] consiste à savoir jusqu'à quel point le programme mis en place a atteint le but principal établi au départ [35]. Ce **modèle rationnel** ou **axé vers les buts**, a été le modèle dominant en évaluation. Ce type d'évaluation vise essentiellement à apprécier les résultats obtenus, sans trop se préoccuper des processus qui interviennent lors de l'implantation du programme [35].*
- *Le deuxième modèle, au contraire, définit l'organisation comme un système complexe et dynamique poursuivant [...] des buts [...] beaucoup plus fluides et instables [...] et constamment redéfinis par les acteurs en présence dans l'organisation et dans l'environnement [35]. Dans ce **modèle systémique**, l'accent est donc davantage placé sur les processus organisationnels mis en place pour assurer le fonctionnement et la survie de l'organisation. Les critères d'évaluation de la performance sont alors établis par rapport à l'utilisation optimale des ressources pour tenir compte de la nature changeante de l'organisation, des multiples acteurs et de leurs points de vue divergents. La difficulté est de déterminer dans un tel contexte ce qui est optimal [35]. La sociologie des organisations propose un modèle d'analyse des enjeux des acteurs dans la situation. Il s'agit d'analyser la place de l'acteur dans le système, pour comprendre les stratégies qu'il va développer afin de garder ou modifier sa position en fonction de celles des autres acteurs présents. L'analyse de ces enjeux permet fréquemment de repérer les zones de pouvoirs et la performance du système. Par exemple, l'analyse institutionnelle repose sur une identification des acteurs et institutions, de leurs rapports entre eux (ou contrats), de leurs valeurs (ou conventions) et des modèles d'action privilégiés [22].*

Modèle axé vers les buts ou modèle systémique ?

Au modèle axé vers les buts correspond une approche d'évaluation basée sur la **démarche hypothético-déductive** utilisée en biologie. Elle privilégie les données quantitatives, l'objectivité, la distance par rapport aux observations, l'approche fragmentaire [35]. Il est *associé à l'évaluation sommative* [35], et plus particulièrement à l'évaluation des résultats.

Au modèle systémique correspond une approche d'évaluation basée sur la **démarche inductive** utilisée en sciences humaines et sociales. Elle privilégie les données qualitatives, la subjectivité, la proximité des milieux d'observation, l'approche holistique [35]. Ce modèle est surtout associé à l'évaluation formative, [qui] *permet, grâce à une rétroaction continue, de réajuster ou de modifier certains éléments du programme pour en améliorer la performance* [35]. Il est également intéressant dans l'évaluation sommative. En effet, *répondre à la question "est-ce que ça marche ?" ne suffit pas, encore faut-il pouvoir répondre aux questions "pourquoi ça marche ?" et "comment ça marche ?". C'est l'enjeu d'une démarche évaluative qui ne soit pas orientée vers la nécessité d'obtenir des résultats quantifiables* [2].

Aucune de ces deux approches n'est meilleure que l'autre : *différentes méthodes sont indiquées pour différentes situations*. Il est donc préférable *d'identifier et d'organiser les preneurs de décision et utilisateurs de l'information produite par l'évaluation* pour définir à partir de leurs besoins la ou les méthodes à utiliser, ou protocole³⁷ d'évaluation [35]. *L'évaluateur devrait être suffisamment souple dans sa démarche pour être en mesure d'adapter ses approches et ses méthodes aux questions envisagées. L'application de méthodes hypothético-déductives nécessite une compréhension nuancée et détaillée tant de l'intervention que des résultats qu'elle entend produire et l'application de mesures qualitatives enrichit et complète les informations recueillies à partir de mesures essentiellement quantitatives* [4]

Nous allons détailler d'abord l'approche systémique, basée sur la méthode sociologique, puis l'approche axée vers les buts, basée sur la méthode épidémiologique.

³⁷ Les Nord-américains utilisent le terme de "*devis*" (en anglais *device*), que nous n'avons pas repris car il désigne en France l'estimation des prix pour un travail demandé.

L'approche systémique du processus et de l'impact

La méthode sociologique, qui dérive de la recherche en sciences humaines et en sciences sociales, est de plus en plus utilisée. Elle est basée sur le recueil de données de nature qualitative obtenue au moyen d'études de cas, d'observations participantes, et d'entretiens semi-directifs. *L'évaluateur vise la compréhension d'un phénomène par sa description plus ou moins exhaustive, cherche avant tout à décrire et à comprendre le fonctionnement du programme, à identifier les informations que désirent obtenir ceux qui sont engagés dans l'intervention et à sélectionner les méthodes de recherche qui conviennent plus particulièrement aux interrogations soulevées [4].*

Elles conviennent particulièrement aux évaluations formatives, où c'est la rétroaction qui est visée. Elle est utilisée dans l'évaluation du processus et de l'implantation. Elle sert également dans l'évaluation d'impact, pour identifier des effets imprévus, et dans l'évaluation de la pertinence du programme.

Les outils utilisés sont : la recherche documentaire (à la recherche d'autres expériences voisines), la lecture des dossiers et archives du programme, l'observation directe des réalisations du programme, les opinions d'experts, de responsables, de personnes concernées par le programme, de bénéficiaires, exprimées lors d'enquêtes ou d'entretiens collectifs ou individuels [31].

*Ces méthodes se réclament d'une rigueur scientifique aussi contraignante que celle des méthodes statistiques. Il faut sortir des approches intuitives et trop personnelles. **Produire des connaissances**, c'est sortir d'une approche trop subjective ou approximative, fondée sur des perceptions et des expériences individuelles ; c'est mettre en place et assurer régulièrement un recueil de données répondant à des règles précises [33].* Quatre types de données peuvent être recueillies : il faut s'entendre sur les valeurs, définir les critères et indicateurs, se constituer un recueil de données d'activités, et recueillir le point de vue du public.

S'entendre sur les valeurs. *La question des valeurs qui sous-tendent le programme est rarement discutée. Or en matière de prévention et de promotion de la santé, un même objectif peut recouvrir des visions très différentes de ce qu'est la santé, la sexualité, l'insertion sociale. [...] Débattre de ces questions est fondamental et devrait être fait au démarrage du programme, mais si ça n'a pas été le cas, la mise en place d'une évaluation obligera à le faire [33].*

Définir des critères qui permettront de décrire l'atteinte des objectifs. Chacun de ces critères sera traduit en **indicateurs** qui pourront s'exprimer

en terme de présence ou d'absence. Il s'agit de décomposer l'objectif principal, ou l'état auquel on souhaite arriver, en étapes ou en états intermédiaires que l'on peut définir et décrire ou mesurer. Les indicateurs [...] serviront à la construction d'un recueil de données. Tous les indicateurs auxquels on peut penser pour évaluer une action ne sont pas forcément accessibles ou d'un intérêt égal. Il faudra trier, et retenir ceux qui sont les plus utiles à l'évaluation et qui ne posent pas trop de problèmes à documenter [33]. On peut se servir d'indicateurs utilisés dans le cadre d'activités similaires déjà évaluées [2], ce qui facilite les comparaisons.

Cinq catégories de mesures de changement sont utilisées dans les études évaluatives en éducation pour la santé :

- 1 – l'acquisition de connaissances,*
- 2 – la variation d'attitudes,*
- 3 – la modification déclarée de comportement,*
- 4 – l'acquisition d'aptitudes à réagir face à une situation,*
- 5 – la modification d'un certain nombre de caractéristiques personnelles [26].*

Encadré 12. Il n'y a pas de "recette" en évaluation de programme.

Dans une approche systémique, on ne peut concevoir un modèle d'évaluation standard, qui puisse être appliqué à tout programme, comme une "recette". Chaque programme est différent, de par ses actions, son contexte, ses objectifs, et il lui faut des critères et indicateurs d'évaluation qui lui soient propres. Les indicateurs standardisés, qui sont intéressants pour les comparaisons, sont souvent peu interprétables pour un programme donné. Ceci est particulièrement vrai pour les programmes d'éducation pour la santé, qui ont leur spécificité. *Ce qui est spécifique, ce sont les objets de l'évaluation due à la nature de l'action d'éducation pour la santé. L'évaluation s'adaptera à cette spécificité par le choix des indicateurs, des critères, des instruments, du recueil d'information, des analyses. Mais la nature de l'évaluation ne change pas [16].*

Pour l'évaluation du processus, il faut se constituer un **recueil de données d'activité** tout au long de l'action. Les activités ne se limitent pas à celles en direction du public visé. Il y a aussi les activités de coordination et concertation avec les financeurs, partenaires et décideurs, et les activités de communication. Le recueil de données fait aussi appel à des indicateurs. Comme pour les objectifs, *on ne peut pas tout recueillir. Il faut se laisser guider par la faisabilité (données observables, faciles à définir, à recueillir, temps de recueil pas trop long) et l'utilité (certaines données ne présentent que peu d'intérêt pour l'évaluation du processus ou des résultats). Chaque activité que l'on veut évaluer demande un recueil de données spécifiques. Il*

ne faut cependant pas oublier les sources d'information "naturellement" disponibles, comme les compte-rendus de réunion, les éléments de communication... [33].

Le recueil des données peut être complété *par le point de vue du public auquel est destiné l'action* [33]. Des entretiens semi-directifs ou des enquêtes de satisfaction, par exemple, peuvent servir de base de données.

Le recueil de données relève de deux approches complémentaires : les enquêtes qualitatives et les enquêtes quantitatives.

Les **enquêtes qualitatives** utilisent des outils tels que :

- **L'entretien individuel** à partir d'une grille d'entretien construite par l'évaluateur en fonction des questions posées et des données qu'il cherche à recueillir. Les entretiens individuels sont arrêtés *lorsqu'on aura atteint ce qu'on appelle la saturation des données ; ceci signifie que même si on ajoutait des répondantes ou répondants, ceux-ci n'apporteraient plus rien de nouveau aux réponses déjà recueillies* [21]. La saturation se produit d'autant plus rapidement que la population étudiée est homogène dans ses caractéristiques et ses comportements, souvent entre quinze et trente entretiens.
- **Les entretiens collectifs et groupes focaux** lors desquels plusieurs partenaires s'expriment à partir de questions préalablement élaborées, en apportant le plus de précisions possibles eu égard aux connaissances de chacun. L'entretien de groupe a l'avantage de permettre l'interaction entre les participants, qui peuvent *construire ainsi un aperçu de l'objet*. Il permet aussi d'apprécier les normes en matière de représentation et de comportements. Le groupe focal diffère de l'entretien collectif par la composition du groupe et par l'objectif de la réunion. Le groupe focal regroupe des personnes diversifiées connaissant bien le sujet et ayant une grande capacité à s'exprimer ; il a pour but de diversifier les points de vue plus qu'aboutir à une construction commune ; il nécessite une animation par un "facilitateur" [21].
- La participation à des **réunions institutionnelles**, apportant des éléments sur le processus de travail du programme à évaluer, ou la lecture de comptes-rendus de réunions et tous **documents** pouvant rendre compte de ce processus.

Encadré 13. Réaliser des entretiens et les analyser.

L'entretien est *un procédé d'investigation scientifique, utilisant un processus de communication verbale, pour recueillir des informations, en relation avec le but fixé* [23]. Il s'appuie sur une grille d'entretien qui saisira les matériaux les plus significatifs par rapport à l'objectif poursuivi par l'évaluation. Pour la conduite d'un entretien, l'enregistrement est parfois utilisé mais, dans la pratique, il n'est pas toujours possible (il est parfois mal accepté, et sa retranscription et son exploitation sont coûteuses).

Une fois les entretiens terminés on vérifie la concordance et l'uniformité des réponses. Ensuite on s'interroge sur la pertinence *d'étudier les interviews par personne ou de rassembler les réponses de tous les enquêtés question par question* [23]. Toutefois, *dans une première étape, les réponses par enquêté doivent toujours être examinées, car elles seules font apparaître les contradictions ou, au contraire, le lien entre certaines questions qui se complètent. Elles permettent également de mieux juger le travail de l'enquêteur. Ce n'est qu'ensuite que l'on aborde, s'il y a lieu, l'étude de toutes les réponses, question par question. Ce tableau reflète mieux l'opinion globale de l'échantillon sur un point précis, il est indispensable pour pratiquer une analyse de contenu générale, ensuite on procède à l'analyse de contenu des propos recueillis.*

L'analyse de contenu, est une technique de recherche pour la description objective, systématique et quantitative, du contenu manifeste des communications, ayant pour but de les interpréter [23] :

- *Objective. – L'Analyse doit procéder selon des règles, obéir à des consignes suffisamment claires et précises pour que des analystes différents, travaillant sur le même contenu, obtiennent les mêmes résultats. Ceci suppose qu'ils se soient mis d'accord sur les aspects à analyser, les catégories à utiliser et la définition opérationnelle de chaque catégorie.*
- *Systématique. – Tout le contenu doit être ordonné et intégré dans les catégories choisies, en fonction du but poursuivi. Des éléments d'information ayant trait à l'objectif ne doivent pas être laissés de côté.*
- *Quantitative. – Le plus souvent, il s'agit de dénombrer des éléments significatifs, de calculer leur fréquence, etc. Mais cette condition n'est pas indispensable et certaines analyses de type qualitatif recherchent les thèmes plus qu'elles ne les mesurent* [23].

Les **enquêtes quantitatives** s'adressent à un plus grand nombre de personnes par l'intermédiaire de questionnaires proposant des questions fermées ou à choix multiples. Ce type de recueil de données permet de repérer des régularités sur un point précis et de dégager des points de vue dominants. Il est plus souvent utilisé dans l'approche épidémiologique. Les questions peuvent porter sur la perception d'une action ou sur ses effets, par exemple le *degré de satisfaction, degré d'atteinte des objectifs* [ou] *la fréquence de comportements, d'attitudes, de savoirs, de représentations ou de croyances* [33].

L'approche qualitative est préférable au questionnaire quand le retour des questionnaires risque de poser problème (population mouvante, public peu à l'aise avec l'écrit) ou quand les domaines explorés sont complexes (les questions ouvertes risquent alors d'avoir des réponses générales ou trop vagues) [33].

Encadré 14. Les règles éthiques des enquêtes

Quel que soit le type d'enquête, il y a des règles éthiques à respecter. Les principales sont : le consentement libre et éclairé des répondants, le droit de se retirer de l'étude sans préjudice, l'anonymat des répondants, la confidentialité des réponses individuelles, et la demande d'autorisation de citer des passages s'il s'agit d'entretiens. Lorsque les répondants sont mineurs, il faut obtenir l'autorisation des parents ou tuteurs [30],[21]. Il convient d'être particulièrement vigilant sur ces règles éthiques lorsqu'on a affaire à une population "captive" telle que des jeunes en milieu scolaire.

L'approche épidémiologique des effets

La méthode épidémiologique³⁸ *pose comme hypothèse l'existence d'une relation entre le programme (variable indépendante) et les effets désirés (variable dépendante). La vérification de cette hypothèse requiert ensuite un protocole de recherche* [35]. Ce protocole décrit les données qui devront être recueillies, leur mode de recueil, et la façon dont elles seront analysées. Il s'agit ici **d'analyse statistique** qui va permettre dans un premier temps de décrire les effets de façon quantitative et dans un second temps de réaliser des tests statistiques reliant ces effets à différents facteurs dont la participation au programme.

³⁸ Seule l'épidémiologie évaluative (qui analyse les effets des interventions) est présentée ici. L'épidémiologie comporte classiquement deux autres domaines qui ne sont pas abordés ici: l'épidémiologie descriptive (qui vise à informer sur une situation sanitaire) et l'épidémiologie analytique (qui met en relation des états de santé avec des facteurs de risque).

Rappelons que tester cette relation entre programme et effets ne peut se faire qu'après une analyse du programme et de son implantation, pour montrer la réalité de l'atteinte des bénéficiaires, une certaine homogénéité ou stabilité des interventions, ainsi que pour définir "l'exposition"³⁹ à ces interventions.

Evaluer veut dire comparer et la comparaison la plus habituelle, pour évaluer une intervention préventive ou thérapeutique consiste en celle de deux groupes : un groupe de sujets bénéficiant de l'intervention à évaluer, ou groupe intervention, et un groupe de sujet n'en bénéficiant pas, le groupe témoin⁴⁰ [5].

*A partir de cette comparaison, on souhaite porter deux jugements : un jugement de signification (est-ce que les résultats dans le groupe [intervention] sont significativement meilleurs que dans le groupe [témoin] ?) et un jugement de causalité (est-ce que la différence de résultats constatée est expliquée par la différence entre les deux interventions ?). Le **jugement de signification** dépend de calculs statistiques. Dans le **jugement de causalité** intervient pour une grande part la méthodologie de l'étude d'évaluation et en particulier son adéquation aux objectifs de l'étude [5].*

*La **validité interne** [de l'évaluation] permet de conclure que l'effet obtenu est réellement dû à la variable indépendante et non à des facteurs ou variables de confusion ("confounding factors"), l'effet de ces dernières ayant été pris en considération et contrôlé dans le [protocole]. La validité interne est garantie lorsqu'on contrôle l'influence des variables concomitantes ou variables de confusion sur l'effet du programme. Une variable de confusion est une variable qui est associée aussi bien à la variable indépendante (le programme) qu'à la variable dépendante (l'effet du programme) [35].*

Exemple de variable de confusion : un programme de cours prénatals est organisé pour améliorer les connaissances et les comportements de santé des femmes durant la grossesse. Ces connaissances et comportements sont mesurés par un test avant les cours (prétest) et après (posttest). Si une série d'émissions de télévision sur la santé des femmes enceintes est diffusée en même temps que les cours, on ne sait pas si la différence observée entre les deux tests est due aux cours prénatals ou aux émissions de télévision. Le facteur "émission de télévision" est un facteur de confusion dans la relation entre le programme (les cours prénatals) et l'effet sur les connaissances et comportements mesuré par les tests.

³⁹ Par analogie avec la liaison d'une maladie avec l'exposition à un facteur de risque, on teste ici un état de santé (ou tout autre effet) en relation à l'exposition à une intervention, qui à l'inverse d'un facteur de risque est supposée avoir un effet protecteur.

⁴⁰ Ce groupe témoin n'est soumis à aucune intervention ou bénéficie de l'intervention habituelle selon le cas.

Lorsqu'on réalise une comparaison, le risque majeur de biais⁴¹ est lié à la non-comparabilité des groupes étudiés, c'est-à-dire au fait que les deux groupes diffèrent sur d'autres points que l'intervention. Deux sont particulièrement importants : le comportement vis-à-vis de la prévention et les caractéristiques pronostiques des deux groupes comparés. [...]. Les solutions proposées pour éviter [les] biais [...] sont de deux ordres : l'expérimentation et le tirage au sort [5].

- L'expérimentation signifie que l'observateur choisit la population à laquelle l'intervention doit être appliquée, décide du moment de cette application, et met en place les moyens d'informations nécessaires à l'évaluation. Un avantage important est d'assurer que la mise en place de l'intervention ne s'accompagne pas d'autres modifications de comportement vis-à-vis de la prévention [5]
- Le tirage au sort [ou randomisation] veut dire ici que l'appartenance au groupe [intervention ou témoin] est déterminé de façon aléatoire. Cette façon de procéder a pour but d'assurer que les deux groupes sont comparables pour tous les facteurs pronostiques (connus ou inconnus) [5].

Il y a trois grands types de protocoles, par degré de validité interne décroissant : expérimental, quasi-expérimental, et pré-expérimental (ou d'observation) [3] (Cf. encadré 15).

Encadré 15. Les protocoles d'études épidémiologiques.

1. Protocole expérimental : c'est l'essai clinique randomisé (*randomized clinical trial*) : l'investigateur exerce un contrôle sur l'intervention et procède à l'allocation des sujets de façon aléatoire [en deux groupes] suivis et comparés en regard de certains indices d'impact. Pour minimiser les biais possibles, la randomisation devrait de préférence s'effectuer à double insu⁴², sans que le sujet et le thérapeute participant à l'étude ne connaissent l'appartenance des sujets randomisés aux groupes constitués [3].

Le tirage au sort peut être individuel (pour une intervention individuelle) ou par groupe (pour une intervention collective). L'unité de tirage au sort est la plus petite unité d'intervention [5].

Ce protocole est le plus valide, mais il pose des problèmes : coût élevé, durée longue, groupes pas toujours représentatifs, perdus de vue,

⁴¹ Biais : erreurs systématiques dans la mesure par rapport aux valeurs réelles.

⁴² Double insu, ou double aveugle : ni les sujets ni l'expérimentateur ne savent l'appartenance des sujets à un des deux groupes (groupe expérimental ou groupe témoin). Simple insu, ou simple aveugle : seuls les sujets ignorent leur appartenance à un des deux groupes.

problèmes éthiques [3],[35]. De plus, il n'est pas toujours possible de répartir les sujets au hasard.

2. Protocole quasi-expérimental : l'intervention est contrôlée par l'investigateur mais la randomisation (le tirage au hasard) est impossible. On distingue les études *ici-ailleurs* et *avant-après* :

* Les **études *ici-ailleurs*** comparent deux groupes géographiquement distincts, un groupe exposé à l'intervention, et un groupe témoin qui ne l'est pas. Les problèmes des protocoles *ici-ailleurs* sont les mêmes que pour le protocole expérimental, ainsi que l'absence de contrôle des variables de confusion non identifiées par le chercheur, et l'absence d'insu⁴³. Ces biais favorisent habituellement le groupe exposé [3]. D'autre part, il faut s'assurer de *l'étanchéité de l'intervention*, car le groupe témoin ne doit pas être concerné par l'intervention étudiée [22].

* Les **études *avant-après*** concernent un groupe observé avant et après l'intervention (pré-test et post-test) [3],[5]. Le problème est un effet de confusion de l'histoire (des événements importants non contrôlés survenant en même temps que l'intervention). Il est possible de faire plusieurs mesures avant et après le programme (séries chronologiques).

Il est possible de combiner les deux méthodes *avant-après* et *ici-ailleurs* : un groupe témoin peut aussi être observé avant et après l'intervention pour prendre en compte l'évolution spontanée hors du programme (effet de maturation, ou effets d'autres événements).

Les protocoles quasi-expérimentaux sont les plus valides lorsque la randomisation n'est pas possible, sous réserve d'une bonne analyse des risques de biais.

3. Protocole pré-expérimental ou d'observation : l'investigateur ne contrôle pas l'intervention et ne peut randomiser. Les résultats de ce type d'études sont à interpréter avec prudence, car le risque de biais est ici très important. On distingue deux types d'études d'observation, qui toutes deux sont rétrospectives⁴⁴ [3] : les études *cas-témoins* et les études transversales.

* **Etudes *cas-témoin*** : deux groupes sont constitués selon un état final et comparés en fonction de leur participation antérieure à l'intervention et à

⁴³ Voir note précédente.

⁴⁴ Les études rétrospectives se situent après l'intervention et s'intéressent à la façon dont la personne a été "exposée" à celle-ci. Elles se différencient des études prospectives présentées dans les protocoles expérimentaux et quasi-expérimentaux où on constitue avant l'intervention des groupes qu'on suit jusqu'à la fin de celle-ci.

d'autres facteurs explicatifs. Les difficultés sont de constituer le groupe témoin et de contrôler les variables de confusion [3].

* **Etudes transversales** : il s'agit d'enquêtes dans une population où on mesure simultanément la participation à l'intervention et un état final [3].

Les études d'observation sont de réalisation rapide, peu coûteuses et ne posent pas de problèmes éthiques, mais leur validité est conditionnée par une bonne analyse des risques de biais.

Les **biais** possibles sont très nombreux dans les études évaluatives. Ils peuvent être dus à l'histoire (des événements autres que le programme), à la maturation (l'évolution spontanée des personnes), aux tests et aux instruments de mesure (qualité et reproductibilité des mesures, régression statistique vers la moyenne), à la sélection des personnes (non comparabilité des groupes au départ ou suite à des abandons). Les facteurs de confusion les plus connus peuvent cependant être neutralisés, soit par le protocole dans le recrutement des groupes étudiés (appariement, stratification), soit au moment de l'analyse par des techniques statistiques (standardisation, analyse multivariée) [35]. Les nombreux obstacles à la validité interne des études évaluatives font qu'il vaut mieux s'y engager avec l'aide d'une personne qualifiée et qu'il ne faut pas trop faire "parler" les chiffres.

Reste la dernière question, une fois qu'une différence entre groupes a été mise en évidence et a été attribuée au programme : ces résultats *peuvent-ils être généralisés à un ensemble d'individus plus grand que celui étudié ?* La **validité externe** permet de généraliser les résultats à une population plus grande ou à d'autres situations ou contextes plus larges [35]. C'est la représentativité de l'échantillon étudié qui permet d'extrapoler les résultats à toute une population [35]. A défaut de tirage au sort, qui est la meilleure méthode, on s'approchera de la représentativité par des critères comme le sexe, l'âge, le lieu de résidence, le niveau socio-éducatif [33].

Encadré 16. Réaliser une enquête par questionnaire et l'analyser.

Les études évaluatives s'appuient sur des enquêtes quantitatives qui *s'adressent à un grand nombre de personnes (le plus souvent plusieurs centaines)* [33]. Elles sont basées sur des questionnaires qui peuvent être remplis par un enquêteur ou auto-administrés. Ils visent à recueillir des données générales sur les personnes, des données sur l'intervention et sur les facteurs de confusion connus, et des données de résultat ou d'impact. Comme on l'a vu, les questionnaires descriptifs peuvent également servir à évaluer les opinions des personnes cibles sur la qualité de l'intervention.

Le questionnaire doit être aussi court que possible : il faut cibler les questions *sur des choses pertinentes, utiles, significatives pour l'évaluation*. Si on utilise un questionnaire déjà existant, il faut obtenir l'autorisation de l'utiliser. Si on construit le questionnaire soi-même, il faut suivre quelques règles : privilégier les réponses fermées (oui/non, échelle numérique ou semi-quantitative⁴⁵, liste de réponses au choix), utiliser des mots simples et non ambigus, formuler des questions affirmatives (sans négation), vérifier qu'il n'y a qu'une seule question par énoncé. Enfin, il est recommandé de tester le questionnaire auprès de quelques personnes du groupe à interroger avant la passation générale [21].

L'analyse des données est d'abord descriptive, et consiste à calculer les distributions de fréquences des réponses, ou des moyennes pour les variables quantitatives [21]. Des analyses croisées ou multivariées permettront d'effectuer les tests statistiques liant les effets observés et l'intervention.

Les points clés de l'évaluation : mesure et appréciation

La validité de l'évaluation repose en premier lieu sur le protocole de l'évaluation, dans le choix des méthodes et des critères. Elle repose ensuite sur la validité des instruments de mesure que sont les indicateurs et la qualité de leur recueil, ce qui pose la question de la validité des déclarations recueillies dans les questionnaires ou par entretiens. Elle repose enfin sur la validité de l'appréciation, donc de la référence qui sert de comparaison.

La validité des mesures a deux grandes composantes, la fiabilité et la validité :

⁴⁵ Par exemple : pas du tout, un peu, beaucoup, tout à fait.

- Une mesure **fiable** est celle qui produit des résultats constants lorsqu'elle est appliquée de façon répétitive. Il s'agit donc ici de la variation due au hasard (random error). Autrement dit, plusieurs mesures faites par des observateurs différents ou à des moments différents doivent donner le même résultat. La fiabilité est une condition nécessaire, mais non suffisante, pour la validité [35].
- La **validité** d'une mesure renvoie à la capacité de celle-ci à mesurer réellement ce qu'elle doit mesurer. En d'autres termes, une mesure valide⁴⁶ représente réellement l'objet de la recherche [35].

Ces questions de validité sont centrales dans l'évaluation des programmes de prévention sanitaire, de promotion de la santé, ou de développement social. *L'état de santé d'une population [...] est une notion très abstraite. Les actions de prévention et de promotion de la santé [...] visent [souvent] des modifications ou des adaptations des comportements [...]. Mais les comportements humains sont complexes et les facteurs qui les influencent sont eux-mêmes complexes. Les comportements ne sont pas seulement influencés par les connaissances, ou par des éléments rationnels. Bien d'autres facteurs, culturels, sociaux, économiques, juridiques et politiques entrent en jeu. D'autre part, la plupart des comportements humains liés à la santé sont difficilement observables, du moins dans leur totalité [...]. Enfin, modifier un comportement ou l'adapter en fonction de données nouvelles prend du temps [et] maintenir cette modification du comportement n'est pas facile [33].*

De plus, il est souvent plus pertinent de s'intéresser aux pratiques des individus, c'est à dire leurs comportements (re)contextualisés dans leur environnement. Cette approche compréhensive oblige à prendre en considération les éléments sociaux et familiaux explicatifs de la stabilité ou des changements des comportements de la personne. En soi, le comportement n'est rien sans l'analyse du contexte dans lequel il se développe, perdure ou se transforme. Ce contexte est composé d'éléments dont les interactions vont plus ou moins déterminer ou permettre les modifications de penser, d'être ou les manières de faire du sujet.

Ces difficultés obligent à interroger systématiquement la validité des indicateurs, parfois complexes (par exemple, la qualité de vie, l'estime de soi), et des modèles explicatifs qui sont utilisés lors d'une évaluation des résultats. Il est donc recommandé de diversifier les indicateurs pour aborder le même concept sous plusieurs angles, d'utiliser des outils déjà validés et

⁴⁶ On peut distinguer la **validité de contenu**, qui consiste, pour la mesure, à représenter toutes les dimensions, tous les aspects du concept qu'elle mesure, la **validité de critère** [qui] consiste à comparer une mesure avec une autre variable extérieure au phénomène étudié, et la **validité de construit** [qui] repose sur la théorie relative au domaine étudié [35].

de ne créer les siens que si aucun outil n'existe déjà, et enfin de "triangler" (Cf. encadré 17).

Encadré 17. Renforcer la validité par la "triangulation".

La "triangulation" est une *barrière de sécurité* qui renforce la crédibilité de l'évaluation. Elle consiste à étudier la concordance ou la discordance entre différentes données, sources de données ou méthodes de collecte de données sur un même phénomène (l'objet d'évaluation). Il y a donc trois triangulations possibles [21] :

- **Triangulation des données** : il s'agit de *comparer un groupe de données avec un autre*, que les sources et les méthodes de collecte soient ou non les mêmes. Par exemple, comparer les changements d'attitude décrits par des usagers avec les objectifs atteints par des intervenants dans un programme semblable ; ou croiser la satisfaction des participants avec leur acquisition de connaissances.
- **Triangulation des sources** : il s'agit de *prendre une même technique [...] mais l'appliquer à des sources de données différentes*. Par exemple, comparer l'appréciation d'usagers et d'intervenants, vus chacun en entretien de groupe, sur le programme.
- **Triangulation de la méthode** : il s'agit d'*appliquer des moyens ou des techniques d'évaluation différents [...] pour produire les mêmes genres d'information [...]*, ce qui permettra de voir si un groupe de données confirme ou non l'ensemble de données produites par l'autre moyen. Par exemple, analyser le degré de satisfaction des usagers par enquête par questionnaire et par entrevue de groupe [21].

Outre les problèmes de mesure qui viennent d'être évoqués, la validité de l'évaluation dépend de la **validité de l'appréciation** qui est donnée, donc de celle des normes qui servent de comparaison. Rappelons que la norme ou standard est la valeur du critère ou de l'indicateur qui est considérée comme acceptable (Cf. page 24).

Les normes peuvent être établies de deux façons : 1) de façon normative, c'est-à-dire généralement à partir du jugement et de l'opinion d'un groupe d'individus bien informés et considérés comme compétents dans le domaine ; ou 2) de façon empirique, c'est-à-dire sur la base d'expériences antérieures provenant soit de la recherche, soit de la comparaison avec d'autres systèmes ou programmes jugés semblables [35].

Les normes peuvent être de différents registres : normes biologiques (par exemple, la durée de la grossesse), normes administratives (par exemple, quatre consultations prénatales pour le suivi d'une grossesse normale), normes professionnelles de qualité (par exemple, recherche d'albumine dans les urines à chaque consultation prénatale), normes définies dans le cadre du programme (par exemple, les objectifs fixés, les activités décidées, les ressources devant être mobilisées).

*Les critères et les normes peuvent être implicites ou explicites. Les critères et normes **implicites** ne sont pas fixés à l'avance, mais plutôt laissés au jugement de l'évaluateur : on fait confiance à sa connaissance du sujet. Les critères et normes **explicites** sont établis d'avance : ils sont alors connus et utilisés de façon uniforme par les évaluateurs [35].*

Evidemment, le problème posé par les critères et les normes est principalement celui de la validité. Ils peuvent être définis par les professionnels, les bénéficiaires, un groupe de pression, les politiques : selon les individus concernés, les critères et les normes risquent d'être définis de façon fort différente [35]. De plus, il n'y a pas de normes pour tout. On peut être amené à examiner des activités [...] pour l'évaluation desquelles il n'existe pas d'indicateurs ni de critères appropriés. Dans ces cas, on posera des questions pertinentes sur l'activité à évaluer [34] : le recours aux normes implicites s'impose, celles-ci devant être discutées par les acteurs du programme.

Chapitre 6 : Les étapes de l'évaluation

L'essentiel :

La mise en œuvre d'une évaluation passe par quatre étapes :

1. Avant de lancer une évaluation, il faut au **préalable** s'assurer qu'elle est pertinente et réalisable.
2. La **préparation** de l'évaluation consiste à mettre en place un comité d'évaluation, à définir (négocier) les questions d'évaluation, les objets de l'évaluation, les critères et les méthodes de recueil des informations, à consigner ces éléments dans un cahier des charges, puis à mobiliser les ressources.
3. La **réalisation** de l'évaluation consiste à collecter les informations et à les analyser, comme indiqué dans le cahier des charges, à en tirer des conclusions et des recommandations, et enfin à rédiger un rapport qui reprend ces différents points.
4. La dernière étape est la **valorisation** des résultats. Elle consiste à diffuser les résultats de l'évaluation, au moyen de supports diversifiés. Il peut être utile de suivre la prise en compte des recommandations.

L'évaluation est un processus lié à l'action, depuis sa conception jusqu'à sa réalisation complète et aux résultats obtenus [1]. L'évaluation doit donc être organisée (préparation), mise en œuvre (réalisation), et ses conclusions doivent être diffusées pour qu'elle soit utilisée (valorisation) [31]. Mais il y a un préalable.

Le préalable : l'évaluation est-elle pertinente et réalisable ?

Avant de planifier une évaluation, il faut s'assurer de sa pertinence et de sa faisabilité. Pour cela, les quatre questions suivantes doivent être posées [21],[30],[31] :

1. *Y a-t-il **un réel besoin** d'évaluer ?* Autrement dit, *un réel besoin de connaissances* à propos d'un programme ou d'un service [21], *une volonté d'amélioration du projet* [31]. Même si l'évaluation est exigée par le bailleur de fond, *il est indispensable qu'il y ait un réel besoin de connaissances et [...] que l'on se pose des questions authentiques* : une évaluation qui n'est pas pertinente est vouée à l'échec [21].
2. *L'organisme est-il prêt au changement ?* Il faut **une ouverture**, *une motivation, une perception et une attitude positive des personnes concernées face à l'évaluation*. Il vaut mieux ne pas évaluer si le personnel *n'est pas prêt à remettre en question ses pratiques ou ses conceptions*⁴⁷ [21].
3. *Est-il possible d'intégrer l'évaluation dans [la] planification du travail ?* Autrement dit, il faut **prévoir la place** qu'occupera l'évaluation [21], et plus généralement pouvoir lui accorder des ressources [31].
4. *Sommes-nous à **un bon moment** pour le faire ?* *On ne doit pas recourir à l'évaluation pour tenter de résoudre un conflit*⁴⁸ *ou un problème à l'intérieur de l'organisme*. Si l'évaluation est faite à la demande du bailleur de fonds, il faut également vérifier qu'il n'y a pas de conflit ouvert avec les partenaires externes [21]. Enfin, l'évaluation ne doit pas intervenir en cas de *situation extraordinaire* ou dans *des périodes de pointe d'activité* [31].

Le fait que l'évaluation a été prévue dès le lancement du programme est un facteur facilitant celle-ci, mais ce n'est pas une condition préalable. Autrement dit, il est toujours possible de faire une évaluation même si rien

⁴⁷ Un temps de préparation et d'explication est nécessaire si des blocages à l'évaluation émergent, afin d'aboutir à une ouverture minimale rendant possible le travail d'évaluation.

⁴⁸ En cas de conflit, c'est un processus de médiation qui doit être mis en place, et non une évaluation qui serait vraisemblablement détournée de son objet.

n'a été prévu au départ. Pour toute évaluation, il faut repartir des intentions de départ, qu'elles aient été écrites ou non.

La préparation : du comité d'évaluation au cahier des charges

Si les réponses aux questions préalables sont positives, on peut passer à la préparation de l'évaluation proprement dite. Cette partie est particulièrement importante puisque *c'est durant cette phase qu'on détermine les enjeux et le contexte du programme ainsi que les questions qui orienteront l'évaluation, que l'on choisit les indicateurs permettant d'y répondre et les techniques de collecte de données* [17]. Elle se décompose en sept étapes [31] :

1. **Mettre en place une instance** responsable de la réalisation de l'évaluation, comité de pilotage ou **comité d'évaluation**. Cette instance doit être indépendante du gestionnaire du projet [31]. Ses membres et son mode de fonctionnement doivent être définis pour garantir le pluralisme, la compétence et la disponibilité pour organiser l'évaluation. *Cinq est un bon nombre pour travailler et ne pas avoir trop de mal à fixer des réunions. Le comité est responsable de l'évaluation, mais ses membres ne sont pas obligatoirement celles et ceux qui feront les opérations.* [21].
2. **Comprendre l'objet** de l'évaluation : le projet à évaluer, son contexte et les hypothèses faites au départ, donc la stratégie, les objectifs, le contenu du programme [31].
3. **Identifier/définir les questions** auxquelles l'évaluation devra répondre, autrement dit QUOI évaluer et POURQUOI évaluer [21]. Il s'agit de définir les objectifs de l'évaluation, puis de traduire les objectifs généraux de l'évaluation en interrogations concernant des faits vérifiables, des processus observables et des grandeurs mesurables. Cette étape est élaborée grâce à la lecture du document de projet, des entretiens avec les commanditaires, et la lecture d'évaluations du même type [31]. Dans la mesure où les questions sont souvent nombreuses, il est préférable de les classer par ordre de priorité [21].
4. **Recenser l'information disponible**, de façon à déterminer les besoins en informations complémentaires [31].
5. **Concevoir une stratégie** d'évaluation : les méthodes, les indicateurs, les techniques de collecte et d'analyse de données. Puis *définir le programme de travail correspondant, les moyens nécessaires, et le budget à prévoir* [31]. Autrement dit COMMENT évaluer et QUI fait quoi dans l'évaluation [21]. *La sélection des informations et arguments à*

rassembler et donc [...] le choix de la méthode d'évaluation à utiliser est certainement la plus grande difficulté dans la conception d'une évaluation. La question à se poser fréquemment est la suivante : est-ce que la démarche d'évaluation recommandée apportera des réponses crédibles aux questions de départ et est-ce qu'elle est pratiquement réalisable dans les limites de temps et de budget imparties ? [31]

6. **Ecrire** les termes de référence que devront suivre les évaluateurs [31], ou le **cahier des charges** [33], ou *devis d'évaluation*⁴⁹ pour les Nord-américains [30],[21]. Une lecture des principales rubriques est dans l'encadré 18.

Encadré 18. Exemple de rubriques du cahier des charges d'une évaluation [31].

- 1- Origine du projet d'évaluation : qui demande l'évaluation et dans quelles circonstances ?
- 2- Objet de l'évaluation : sur quoi porte l'évaluation, et éventuellement ses limites (ce qui est hors évaluation) ?
- 3- Description schématique du projet à évaluer : objectifs, résultats attendus, indicateurs de résultats, partenaires, budget, activités, calendrier.
- 4- Motifs de l'évaluation : pourquoi veut-on évaluer ? Quelles sont les principales questions qui se posent, et éventuellement les hypothèses à valider ?
- 5- Les questions précises posées à l'évaluation, les critères, les résultats concrets attendus de l'évaluation.
- 6- Méthode et plan de travail : approche générale de l'évaluation, informations à collecter, analyses à effectuer, indicateurs à collecter, calendrier de réalisation, quels documents seront produits, quand et pour qui, et quelle(s) restitution(s).
- 7- Dispositif d'évaluation : composition de l'équipe, profil de l'évaluateur externe à recruter si nécessaire, moyens complémentaires à réaliser, calendrier, organisation du travail et répartition des tâches, dispositif de suivi et de compte-rendu progressif.
- 8- Eventuellement, le plan du rapport final.
- 9- Si un évaluateur externe est requis, *le prix, les modalités de facture et de règlement*. Il est aussi précisé *à qui l'évaluateur rendra compte, à qui appartiennent les données et les rapports produits, et comment ils seront utilisés [33].*

⁴⁹ Nous ne recommandons pas ce terme qui est ambigu en France.

7. **Mobiliser les moyens.** Le **calendrier** de l'évaluation *doit être le plus court possible* pour assurer la pertinence de l'évaluation car les préoccupations des décideurs, des acteurs et le contexte se modifient avec le temps [31].

Le choix de l'équipe d'évaluation est primordial, et notamment l'évaluateur externe : il doit être plus animateur et pédagogue qu'inquisiteur. Il devra pouvoir apporter des garanties d'objectivité, travailler en toute transparence, avoir une capacité d'écoute développée, un esprit ouvert, critique mais constructif et avoir des connaissances dans le domaine étudié [31].

Le plus tôt est le mieux pour embaucher une personne spécialisée en évaluation : juste après la formulation du cahier des charges. Si cette personne intervient seulement au moment de l'analyse des données, il y a deux risques : d'une part qu'elle ne soit pas suffisamment imprégnée de la démarche du programme, et d'autre part qu'elle ne puisse pas tirer grand chose des données si les instruments ne sont pas adaptés aux analyses à faire et aux objectifs de l'évaluation [21].

Dans une démarche participative, **la négociation** accompagne la rédaction du cahier des charges de l'évaluation. Elle s'effectue d'abord *aux différents paliers de l'organisme concerné* (conseil d'administration, personnel, membres, usagers). Il s'agit ensuite d'une *négociation externe des formulations acceptées à l'interne*. En retour, il faudra *soumettre à l'interne les résultats de cette négociation*, et continuer ce processus jusqu'à un accord sur le protocole d'évaluation. [21].

Les groupes communautaires menant des actions ne disposent souvent *pas d'une grande marge de manœuvre pour négocier [leurs] conditions d'évaluation* avec le bailleur de fonds [37]. Néanmoins, cette étape, qui a pour but d'aboutir à une entente satisfaisante pour les deux parties ne doit pas être négligée.

Quelques conseils peuvent leur être donnés pour préparer et accompagner cette négociation (Cf. encadré 19).

Le cadre de la négociation n'est pas toujours clairement défini. *Il prend parfois la forme d'un comité de travail plutôt que l'aspect d'une séance formelle de négociation.* Cette réunion doit donc être bien préparée, en ce qui concerne *l'objet de la négociation, c'est-à-dire ce sur quoi porte la négociation* (comme l'objet d'évaluation, les critères d'appréciation) et en ce qui concerne *le processus de négociation, c'est-à-dire comment l'objet est négocié* (comme les attitudes, la tactique de négociation).

En général, la négociation comporte une *étape initiale*, qui vise surtout à *un échange d'informations*, puis une *étape d'émondage*⁵⁰ qui vise à identifier une zone d'entente avec l'autre partie, et une *étape finale* [qui] conduit à la conclusion d'une entente [30].

Encadré 19. Conseils aux groupes communautaires pour négocier avec le bailleur de fonds les conditions d'une évaluation [37].

1. *Passez en revue tout ce qu'a dit [et écrit] le bailleur de fonds pour expliquer pourquoi il finance votre groupe et pourquoi il demande une évaluation.*
2. *Comparez toutes ces données avec les descriptions que vous-mêmes avez données du travail que vous vous proposez de faire. Relevez les points de divergence et de convergence.*
3. *Rédigez votre propre version de l'affaire : pourquoi, selon vous, avez-vous reçu une subvention et pour quelles raisons vous demande-t-on une évaluation ?*
4. *Si vous le jugez utile [...], envoyez cette version au représentant du bailleur de fonds [...]; sollicitez son accord ou demandez-lui des précisions.*
5. *Demandez au bailleur de fonds de préciser ce qu'il attend du processus d'évaluation en tant que tel : exigences, objets d'évaluation, période, critères...*
6. *Soyez énergiques et sûrs de vous. N'oubliez pas que l'évaluation doit d'abord et avant tout répondre aux besoins de votre groupe.*

La réalisation : du recueil des données au rapport final

Réaliser l'évaluation, c'est faire ce que dit notre projet d'évaluation, [même si] certaines parties seront ajustées. Les personnes qui réalisent l'évaluation ont besoin d'une *trousse de travail personnalisée* qui comporte le projet d'évaluation, un contrat précisant leur mission, la déontologie qu'ils doivent respecter, un calendrier des opérations à réaliser, les personnes auxquelles ils peuvent se référer, et des éléments matériels (documents, magnétophone, cassettes, piles, clés, cahiers...) [21].

⁵⁰ Au sens propre, c'est l'élagage des arbres, le tri et le nettoyage des graines. Dans le cadre d'une négociation, c'est le tri parmi les revendications, la modification des demandes initiales et la présentation de contre-propositions.

La réalisation de l'évaluation passe par huit étapes [31] :

1. **Exploitation des données déjà disponibles** (données essentiellement documentaires) et la préparation de la collecte de données nouvelles [31].
2. **Collecte des données** : préparation des instruments d'enquête ou d'entretiens, et leur mise en œuvre [31]. Une attention particulière doit être accordée au respect de l'anonymat des personnes enquêtées et à la confidentialité des données [33],[37].
3. **Analyse**, en multipliant les angles d'approche, en croisant les points de vue et les sources d'information, pour comprendre les logiques d'acteurs [31].
4. **Argumentaire et conclusions**. *Une bonne transparence dans la présentation de l'argumentaire est importante puisque l'objectivité se vérifie par le fait que l'argumentation peut être reconstruite par d'autres que les auteurs de l'évaluation qui doivent pouvoir parvenir aux mêmes conclusions. Les limitations et les forces des conclusions doivent être clairement listées. Il peut arriver [...] qu'il s'avère nécessaire d'obtenir des informations complémentaires en reprenant des travaux de collecte ou d'analyse. Les conclusions doivent être pertinentes par rapport à l'aide à la décision, et en particulier doivent correspondre aux questions posées au début de l'évaluation* [31].
5. **Recommandations**. *Une séparation doit être faite entre conclusions et recommandations. Les recommandations sont des acquis à conforter, des lacunes à combler, des orientations nouvelles à suivre. L'évaluation ne doit pas aller jusqu'à la programmation détaillée de nouveaux objectifs et tâches, qui doit être laissée à ceux qui conçoivent les projets* [31].
6. **Restitution du rapport provisoire**. *Une évaluation doit se donner le temps de la restitution. La démarche suivie [par les évaluateurs] doit être compréhensible [...] par n'importe quel lecteur. La présentation doit bien distinguer le constat de la réalité d'une part et la formulation d'un jugement d'autre part. Sur la base du rapport provisoire, il y a lieu de donner à toutes les parties en présence [...] la possibilité de faire valoir leur point de vue, si possible lors de débats contradictoires largement ouverts. Il peut arriver qu'à l'issue de confrontations les évaluateurs souhaitent reprendre tout ou partie de leur argumentaire et de leurs conclusions* [31].

7. **Le rapport final** doit traduire le double caractère descriptif et analytique de l'évaluation. Les résultats du travail de l'équipe d'évaluation peuvent être complétés par des avis, éventuellement divergents, exprimés lors de la confrontation. Il [doit] comporter un résumé et toutes les données utiles en annexe. Un exemple de plan est donné dans l'encadré 20.

Encadré 20. Exemple de plan d'un rapport d'évaluation [31].

- 1- Un résumé analytique de 1 à 3 pages reprenant tous les chapitres.
- 2- Des éléments d'information suffisants décrivant le projet évalué.
- 3- Un rappel du contexte de l'évaluation, du commanditaire, de ses objectifs (les questions auxquelles l'évaluation devait répondre).
- 4- Un exposé méthodologique rappelant les choix opérés, les techniques utilisées et les étapes du travail d'évaluation.
- 5- Le constat. Un compte-rendu des études réalisées [et] des développements informatifs à partir des données recueillies et des observations effectuées.
- 6- Les conclusions. Des raisonnements bien identifiés permettant de formuler des conclusions par rapport aux critères retenus pour l'évaluation. Bien expliquer les normes utilisées à des fins de comparaison pour juger des niveaux d'atteinte des objectifs.
- 7- Les recommandations, répondant strictement aux demandes adressées aux évaluateurs dans les termes de référence et clairement reliées aux conclusions.
- 8- Les annexes : données de cadrage, liste des personnes rencontrées, instruments de collecte de l'information, rappel des termes de référence...

L'évaluation se termine par un retour sur la démarche d'évaluation ou méta-évaluation. Cette étape consiste à **évaluer l'évaluation**, son déroulement, les événements qui ont eu lieu, les leçons à tirer pour la prochaine évaluation que l'on pourrait avoir à faire, mais aussi les bonnes idées à conserver. Cela peut se faire par un court questionnaire adressé à tous les participants, ou par une entrevue de groupe [21].

La valorisation : diffusion des résultats et suivi post-évaluation

Cette phase est importante car elle permet finalement à l'évaluation de trouver son utilité. Elle comporte deux démarches : la diffusion des résultats et la prise en compte des recommandations [31].

***Diffuser** n'est pas neutre. La décision de diffuser peu ou prou les résultats d'une évaluation conditionnera dès lors la portée et les retombées de celle-ci dans l'environnement social. [Elle] peut avoir un impact relativement important sur le programme en lui-même [...] mais aussi sur les intervenants, les bénéficiaires, les décideurs, les financeurs [2]. L'évaluateur doit donc planifier adéquatement la phase de diffusion des résultats et y consacrer le temps et les efforts nécessaires, même si cela ne garantit [...] pas leur utilisation éventuelle [4].*

La diffusion des résultats se fait sous diverses formes : rapport intégral, résumé analytique, article de revue, publication, réunions, séminaires, ateliers de travail. En particulier, diverses réunions sont souvent organisées avec les responsables des projets pour discuter de la mise en œuvre des recommandations et de l'utilisation des résultats pour la conception des actions futures [31].

Plusieurs facteurs sont à considérer dans la communication des résultats :

1. Il est important que tous les acteurs concernés par l'évaluation aient **accès à la totalité du rapport**. Pour une lecture plus facile, une **synthèse** peut être proposée. Il peut être dangereux de rédiger des rapports différents pour chaque catégorie d'acteurs, car ils risquent de ne pas percevoir la globalité de l'analyse et des recommandations.
2. *La valence⁵¹ de l'information* : mettre l'accent sur les aspects négatifs déclenche des réticences. Il vaut mieux présenter *dans un premier temps les aspects positifs de l'intervention* et ensuite *les éléments qu'il faudra améliorer à court et long terme* [4].
3. Le rapport final d'évaluation doit être rédigé de façon à être **accessible** par l'ensemble des acteurs concernés. Il vaut mieux *éviter l'utilisation d'un jargon technique* même s'il permettrait de renforcer la crédibilité de l'évaluation. *Le rapport d'évaluation ne devrait jamais contenir de surprises désagréables majeures* qui prennent les intervenants au dépourvu, risquent de les menacer, et de se retourner contre l'évaluateur et son rapport [4].

⁵¹ La valence indique la puissance d'attraction (valence positive) ou de répulsion (valence négative) d'un objet.

4. **D'autres supports** de communication que le seul rapport et son résumé peuvent être utilisés (Cf. encadré 21) pour toucher des publics plus larges et divers.

Encadré 21. Diversifier les supports d'information sur l'évaluation.

- Les résultats de l'évaluation peuvent intéresser d'autres organismes qui réalisent des actions semblables. Il peut donc être intéressant de diffuser le rapport ou le résumé à l'extérieur par une publication ou un article, et de communiquer dans des colloques et congrès [21].
- Des panneaux d'affichage avec des explications, des graphiques, des photos, des extraits d'interview, permettent aussi de communiquer avec la population concernée, surtout s'ils sont sur un lieu de passage ou s'ils sont exposés lors d'un événement spécial. Cela permet également de valoriser les personnes ayant participé à l'action, et d'inscrire le programme dans la vie de la cité [33].
- Les media locaux (presse, radio, télévision) peuvent aussi être sollicités. *Mais il ne suffit pas de leur envoyer le rapport d'activité ou le résumé. Rencontrez-les, invitez-les sur le terrain, faites leur rencontrer des participants.* Ces supports sont en général valorisants pour les financeurs et commanditaires et pour le public destinataire. *Mais attention aux effets pervers [d']un mauvais article ou [d']une rupture d'anonymat [33].*

La prise en compte des recommandations. *Le temps de l'évaluation (et de sa conclusion) doit être bien séparé du temps de la décision. Evaluation et décision sont en effet deux moments et deux responsabilités distinctes. [...] La prise en compte des recommandations de l'évaluation dépend des décideurs [31].* L'instance d'évaluation dispose cependant de cinq moyens pour favoriser la prise en compte des résultats et recommandations de l'évaluation (Cf. encadré 22).

Encadré 22. Comment favoriser la prise en compte des résultats et recommandations de l'évaluation [4].

- 1- *Les recommandations doivent suggérer les changements susceptibles d'améliorer le programme étudié. Elles doivent être réalistes, tenir compte des contraintes de l'organisme, explicites [sur] la démarche de leur implantation, et décrire les coûts et ressources inhérents à leur adoption.*
- 2- *La participation du personnel : si elle intervient à chacune des étapes, elle permet à l'évaluateur de respecter les attentes de ce personnel et encourage [...] intervenants et gestionnaires à donner suite à l'étude. Si elle n'a pas été effective au départ, leur participation doit intervenir le moment venu d'interpréter les résultats et de formuler ses recommandations.*
- 3- *Le synchronisme du rapport : il va de soi que l'évaluation doit être conclue avant que les décisions concernant les changements au programme ne soient prises. L'évaluateur a parfois avantage à diffuser ses résultats au fur et à mesure qu'ils sont disponibles.*
- 4- *La validité de l'étude : la pureté méthodologique [de l'évaluation] ne constitue pas en soi une condition suffisante pour assurer l'utilisation de ses résultats. Toutefois, ceci n'implique surtout pas que l'évaluateur opte pour le laxisme et diminue la rigueur de son approche méthodologique.*
- 5- *Le suivi de l'évaluation : il se peut fort bien que l'organisme décide de ne pas modifier son programme. Il est souvent utile de faire le point après quelques temps de la prise en compte des résultats de l'évaluation, même si ce temps n'a pas été prévu par le commanditaire. Il est parfois nécessaire que quelques intervenants servent de catalyseurs afin d'accélérer le processus de changement [4].*

La prise en compte des recommandations est le but de toute démarche d'évaluation. C'est à ce moment qu'on peut finalement juger de l'utilité de l'évaluation.

Une évaluation est véritablement utile si ses recommandations sont prises en compte. **Pour qu'une évaluation soit utile**, elle doit être *crédible (digne de confiance), pertinente et accessible* [31] :

- **La crédibilité** s'appuie d'abord sur la **validité**, donc la **rigueur méthodologique** : qualité et transparence des méthodes employées, qualité des mesures et des relations causales établies, reproductibilité des conclusions, séparation des conclusions et des recommandations dans le rapport. Elle dépend aussi de **l'adhésion** des personnes au processus d'évaluation, donc de leur **confiance**. Plusieurs facteurs y contribuent : explicitation des objectifs et des questions d'évaluation, modestie (évaluation ciblée sur quelques questions principales), caractère participatif et négocié de l'évaluation, "zone d'autonomie" de l'évaluation, qualités des évaluateurs (compétence, impartialité, indépendance par rapport aux gestionnaires du programme).
- **La pertinence** de l'évaluation signifie que l'évaluation **répond bien** aux questions qu'on se pose sur le programme. Plusieurs facteurs y contribuent : questions posées à l'évaluation correspondant aux préoccupations des responsables du programme, du commanditaire et des autres partenaires du programme, évaluation réalisée à un moment opportun du programme, participation des acteurs à toutes les phases méthodologique, souplesse de la démarche pour adapter les approches et les méthodes aux questions et à l'environnement, résultats disponibles à temps pour être pris en compte au moment de la décision.
- **L'accessibilité** de l'évaluation est la facilité avec laquelle on peut se l'approprier. Elle pose essentiellement la question de la **communication**. Il faut y consacrer du temps et des efforts : *l'évaluateur joue le rôle d'un agent facilitateur et de pédagogue dans l'utilisation des résultats*. En effet, *l'information issue de l'évaluation [doit être] considérée comme un outil de négociation entre des intérêts multiples et non comme une vérité absolue* [11]. Plusieurs facteurs contribuent à l'accessibilité. Un rapport structuré, clair, concis présentant de façon assez complète les arguments, distinguant bien les hypothèses, les observations et les appréciations, y contribue [31]. Des recommandations bien identifiées, réalistes par rapport aux contraintes de l'organisme et à l'environnement du programme, et précises sur leurs modalités d'application et leur coût [4]. La diversification des modes de communication, comme la restitution orale, ou d'autres supports.

Encadré 23. Pour réussir l'évaluation.

Les facteurs de réussite, ou critères de qualité de l'évaluation ont été vus tout au long de ce document, à chaque étape. Voici les 5 points clés :

- **L'évaluation est prévue** le plus tôt possible, dès la phase d'élaboration du programme. Il s'agit alors de *convenir d'une stratégie d'évaluation*, plus que de définir complètement le plan d'évaluation, à cause des *modifications et réajustements qu'on apportera inévitablement au programme* [4].
- **Les conditions préalables sont remplies** : réel besoin de connaissance, ouverture par rapport au changement, ressources disponibles pour la réaliser, moment opportun pour la faire en dehors d'un conflit ouvert ou d'une situation extraordinaire.
- **Les enjeux sont pris en compte dans l'organisation** : l'évaluation dispose d'une *zone d'autonomie* [14], s'appuie sur un travail de groupe (comité d'évaluation) [30], et est *participative et négociée* [21].
- **Un processus planifié**. *L'évaluation doit suivre un plan [qui] indique les étapes du processus d'évaluation et la façon d'y participer, les sources et le type d'information à recueillir, et les méthodes à suivre. Il faut préparer [le recueil des données] et éviter d'improviser. Les données doivent donner lieu à une appréciation. Le processus d'évaluation se termine par un rapport qui [...] contient des recommandations d'amélioration* [30].
- **Respect des principes éthiques** : *pas de risque de nuisance ; transparence [...] en informant les acteurs et les participants* [1] ; *règles de confidentialité et de déontologie régissant toute recherche impliquant des sujets humains* [30].

En conclusion, ***une évaluation n'a de sens que si [...] elle produit du changement tout en produisant de la connaissance*** [10].

Conclusion : pour un langage commun en évaluation

Issue des pratiques des acteurs de terrain, mais utilisant des méthodes issues de la recherche, à la jonction des méthodes quantitatives et qualitatives, l'évaluation est un concept aux contours multiples, qui regroupe diverses réalités, selon les questions qui lui sont posées. N'existant de manière formalisée que depuis quelques décennies, il est normal que cette discipline jeune soit foisonnante : chaque auteur ou presque s'est forgé ses définitions et ses classifications. Cette analyse de la littérature montre que, malgré cette diversité (et parfois même des contradictions), il est possible de démythifier l'évaluation, de la structurer et d'adopter un langage commun.

Démythifier l'évaluation

Tantôt repoussoir, tantôt objet du désir, l'évaluation n'est pas un outil "magique". Elle ne va pas du jour au lendemain décréter que tel type d'intervention est "bon" et doit être dupliqué partout, et que tel autre est "mauvais" et doit être abandonné. L'évaluation d'un programme va aboutir à des appréciations nuancées (points forts, points faibles) pour un programme donné, dans son contexte. Ses conclusions pourront, dans certaines conditions, être extrapolées ou comparées à d'autres interventions.

L'évaluation est un processus qui associe un recueil d'informations systématiques et diversifiées selon les points de vue, une appréciation posée sur la base de critères, d'indicateurs et de normes, et qui se termine par des recommandations en réponse aux questions qui sont posées en vue de décisions concernant un programme.

Elle est un processus, à la fois politique et technique, mené dans une logique de changement organisé. Elle s'inscrit soit en cours de programme dans une rétroaction rapide pour le pilotage (évaluation formative), soit en fin de phase ou en fin de programme dans le but d'analyser ses effets (évaluation sommative).

Cependant, cette logique d'amélioration n'est pas la seule en jeu dans un programme. Les acteurs ont des préoccupations différentes, ils ont parfois des buts officieux, voire cachés sur ce qu'il convient de faire de l'évaluation. L'évaluation, tout comme le programme lui-même, s'inscrit donc dans des logiques sociales et des stratégies d'acteurs. Ces enjeux politiques et les risques de manipulation qu'ils comportent, peuvent entraîner des résistances. La première crainte déclenchée par l'évaluation est celle du

contrôle. Il importe donc de différencier l'évaluation du contrôle. En ce sens, la question de "qui évalue" est particulièrement importante.

Néanmoins, l'évaluation reste une utopie possible, si elle s'inscrit dans une zone d'autonomie, si elle est transparente dans ses objectifs, ses critères et ses méthodes, si les acteurs en ont négocié les termes et s'ils sont impliqués à toutes les étapes (que l'évaluation soit interne ou externe).

L'évaluation s'applique sur tout ou partie des composantes d'un programme, et sur tout ou partie des objets de celui-ci. Les cinq objets ou niveaux d'un programme sont sa structure (ses ressources), ses activités (ou processus), les populations-cibles à qui elles sont destinées (ou l'implantation), ses résultats (ou objectifs) et son impact (ou conséquences). A chaque niveau correspond un type d'évaluation au moins. Selon l'objectif formatif (où on s'intéresse surtout aux efforts des acteurs) ou sommatif (où on s'intéresse surtout aux effets dans les groupes-cibles), plusieurs niveaux du programme peuvent être concernés par une évaluation.

Structurer l'évaluation

Les méthodes de l'évaluation des programmes dérivent des méthodes de la recherche épidémiologique et de la recherche en sciences sociales. Chacun de ces champs correspond à un modèle de conception du programme : modèle rationnel orienté vers les buts pour l'approche épidémiologique, modèle complexe ou systémique pour l'approche sociologique. Ces deux approches doivent souvent être combinées pour apporter des réponses pertinentes aux questions posées lors d'une évaluation.

Le recueil des données fait appel à la recherche documentaire, à l'observation, aux entretiens individuels ou de groupe, aux enquêtes quantitatives. Chaque méthode a ses exigences de qualité, le protocole décrivant les méthodes à utiliser à chaque étape est un élément de référence pour assurer la validité de l'évaluation. La triangulation, qui permet de comparer des informations issues de méthodes, de sources ou de données différentes, est une façon de renforcer la validité des données.

L'appréciation est basée sur la comparaison des données à l'aide des références que sont les critères et les normes. En l'absence de normes explicites dans de nombreux domaines des programmes de santé et du social, les normes sont implicites. L'appréciation portée sur un programme lors d'une évaluation n'est donc pas objective (car elle n'est ni neutre ni désengagée) sans être pour autant subjective (puisque'elle se fonde sur des informations systématiquement recueillies avec une méthode objective). Il ne s'agit pas d'un jugement "couperet", mais d'une appréciation sujette à

discussion, à élaboration, à explication dans un but de compréhension et d'amélioration.

Comme le programme lui-même, une évaluation se planifie, se met en œuvre et... s'évalue. Une évaluation est peu utile, et peut même être dangereuse, dans un contexte conflictuel ou bloqué. Il faut donc vérifier au préalable qu'elle est pertinente et faisable avant de la planifier.

L'évaluation standardisée, "prêt-à-porter", n'existe pas. Il ne peut y avoir que du "sur-mesure" : l'évaluation se construit, ce qui suppose un projet, des "objets" et une méthode. On ne peut pas se contenter de "demander une évaluation" d'un programme, sans spécifier en vue de quelle(s) décision(s) elle doit être faite, sans déterminer les questions prioritaires qui sont posées, sans reprendre les intentions de départ (c'est à dire les documents de planification), et sans disposer de données sur le déroulement du programme (rapports ou relevés d'activité...). Ces différents points doivent être vus collégialement par un comité d'évaluation qui va assurer le pilotage de l'évaluation. La rédaction d'un cahier des charges indique les points fondamentaux de l'évaluation et décrit son déroulement.

La réalisation de l'évaluation consiste essentiellement à faire ce qui a été décidé et à rendre compte des informations recueillies, des appréciations, et des recommandations proposées sous forme d'un rapport d'évaluation. D'autres modes de communication que le rapport écrit sont possibles et souhaitables. Du temps doit être consacré à la valorisation des conclusions de l'évaluation. C'est la prise en compte des recommandations qui permet à l'évaluation de trouver son utilité.

Adopter un langage commun

L'évaluation suppose un partenariat effectif et une bonne communication entre évaluateurs, acteurs de terrain, financeurs et décideurs, puisqu'elle est basée sur une diversification des points de vue et nécessite leur clarification. Ceci n'est possible que si tous adoptent le même vocabulaire. C'est l'enjeu principal de ce document que de définir un langage commun de l'évaluation. Les termes principaux vus et expliqués tout au long de ce document sont regroupés dans un lexique-index (Cf. annexe 1).

Quelques conseils pour la pratique...

L'évaluation est souvent demandée par le bailleur de fonds. Ce n'est pas une catastrophe mais une opportunité pour prendre du recul par rapport à ses actions et capitaliser son expérience.

L'évaluation interne permet aux acteurs eux-mêmes de répondre à de nombreuses questions suscitées par les programmes qu'ils mènent. Il est toujours possible de solliciter une aide méthodologique en cas de difficulté. Voici quelques conseils pour "passer à l'acte" d'évaluation :

- L'évaluation permet de s'interroger sur ses activités, sur ses actions et souvent aide à orienter ses décisions. Elle a une dimension sociale ou démocratique, en libérant la parole des acteurs.
- L'évaluation doit être comprise comme une activité au même titre que d'autres activités d'un programme. Comme les autres, elle doit être pensée, programmée et financée. Sa position intermédiaire ou finale ne doit pas laisser penser qu'elle n'est pas nécessaire ou primordiale.
- Il n'y a pas d'évaluation possible s'il n'y a pas de projet. Il faut donc repartir des intentions de départ, des objectifs poursuivis, des activités mises en places et des moyens mobilisés. Un programme est un acte social organisé qui dit quelque chose sur ses auteurs et son environnement.
- L'évaluation en cours, ou évaluation continue, permet de prendre du recul, de faire un "arrêt sur image". Cette position réflexive par rapport à l'action permet d'arrêter l'activisme pour un temps. Souvent, les acteurs des programmes adoptent déjà cette position, mais de manière rapide, parcellaire, et informelle. Le faire de façon méthodique, systématique et transparente coûte un peu plus en temps, mais peut éviter que programme ne "dérape", ce qui peut représenter une économie substantielle.
- L'évaluation ne s'improvise pas, elle fait appel à des méthodes et techniques qu'il est nécessaire de s'approprier par un apprentissage.
- Elle poursuit des objectifs qui s'inscrivent dans un cadre conceptuel. Il est important, dans une démarche d'évaluation, de s'attarder sur les questions éthiques qu'elle soulève et de s'entendre sur le sens de cette démarche.
- Enfin, l'adhésion et l'accord de toutes les personnes concernées sont le minimum nécessaire. Leur participation active est encore plus souhaitable.

ANNEXE 1 : LEXIQUE - INDEX

Note importante

Ce lexique est constitué des définitions que nous avons retenues suite à l'analyse bibliographique. D'autres définitions peuvent exister, parfois dans un sens différent.

Nous avons indiqué les termes que nous considérons comme équivalents, ainsi que les termes qu'il est important de différencier.

Les termes déconseillés n'apparaissent pas ici.

Nous vous invitons à vous reporter au texte pour plus d'explications ou des exemples. La page étant indiquée, ce lexique peut aussi être utilisé comme index.

Analyse de contenu (p. 60). *Technique de recherche pour la description objective, systématique et quantitative, du contenu manifeste des communications, ayant pour but de les interpréter* [23]. Elle est utilisée pour analyser les entretiens individuels ou de groupe, ainsi que les questions ouvertes (à réponse libre).

- Différencier de : analyse statistique.

Analyse de situation (p. 17). *Evaluation systématique de la nature, de la profondeur et de l'étendue d'un problème et des besoins en vue d'établir des priorités et d'implanter [...] un programme* [30].

- Termes voisins : **état des lieux, diagnostic de situation, étude de milieu, analyse des besoins, évaluation des besoins, analyse préliminaire, év. prospective, év. a priori, év. ex-ante, év. pré-décision, év. de la planification, év. de la faisabilité.**
- Différencier de : évaluation post-décision.

Analyse statistique (p. 61). *Elaboration d'informations numériques ou catégorielles en vue de décrire des phénomènes, mettre en évidence des relations entre variables ou généraliser les résultats* [31]. Elle est utilisée pour analyser les questionnaires ou les mesures dans les études quantitatives.

- Différencier de : analyse de contenu.

Audit (p. 37). *Contrôle de légalité et de conformité par rapport aux normes de fonctionnement définies par les lois et les règlements, directives, budgets, circulaires, notes internes* [31].

- Terme voisin : **contrôle.**
- Différencier de : évaluation.

Autoévaluation (p. 38). Modalité particulière de l'évaluation interne pratiquée par des "bénéficiaires" de

l'action ou leurs représentants [...] selon leurs propres critères [19]. Voir aussi évaluation interne.

Avant-après (étude) (p. 64). Catégorie de protocole quasi-expérimental où un ou plusieurs groupe(s) est(sont) étudié(s) avant et après l'intervention (pré-test et post-test).

- Différencier de : ici-ailleurs (étude).

Biais (p. 63-65). Erreur systématique à l'origine d'un écart entre le résultat établi par une étude (ou une mesure) et la réalité. Voir aussi validité des mesures.

Cahier des charges (d'une évaluation) (p. 73). *Ensemble des questions auxquelles l'évaluation devra permettre de répondre et des contraintes que l'évaluateur devra respecter* [31].

- Terme voisin : **termes de référence.**
- Différencier de : protocole.

Cohérence (p. 48). Voir évaluation de la cohérence.

Contrôle (p. 29-30). Processus hiérarchique visant à vérifier que les dépenses, les calendriers et les activités sont réalisés en conformité avec les normes et les plans établis et à sanctionner si ce n'est pas le cas.

- Terme voisin : **audit.**
- Différencier de : évaluation.

Couverture (p. 43). Voir évaluation d'implantation.

Critère (p. 24 & 57). *Un point de vue qui permet de porter un jugement d'appréciation sur un objet* [31]. *Une caractéristique observable [...] correspondant aux éléments du programme évalué* [35].

- Différencier de : indicateur, norme.

Efficacité (Anglais : *effectiveness*) (p. 43, 50). Degré d'atteinte des objectifs dans la population visée par le programme.

- Termes voisins : **efficacité réelle, efficacité populationnelle.**
- Différencier de : efficience, utilité.

Efficience (Anglais : *efficiency*) (p. 50). Evaluation économique qui exprime le rapport des résultats d'un programme ou d'une action de santé aux moyens engagés (ressources humaines, financières et autres, processus et technologie sanitaires, temps) [34].

- Terme voisin : **rendement.**
- Différencier de : efficacité, utilité.

Evaluation (p. 22). Processus qui associe un recueil d'informations systématiques et diversifiées selon les points de vue, une appréciation posée sur la base de critères, d'indicateurs et de normes, et qui se termine par des recommandations en réponse aux questions qui sont posées en vue de décisions concernant une intervention. Voir d'autres définitions de l'évaluation de programme dans l'encadré page suivante.

- Différencier de : audit, contrôle, suivi, supervision.

Evaluation externe (p. 37). Evaluation où l'évaluateur est choisi en dehors de la structure responsable du projet [31].

- Terme voisin : **allo-évaluation.**
- Différencier de : évaluation interne.

Evaluation formative (p. 20, 47). Elle favorise une meilleure connaissance de l'action, permet une prise de conscience critique du travail réalisé, fournit des repères pour réajuster le tir et maintenir le cap : elle est apprentissage dans l'action et action autocorrectrice [30].

- Termes voisins : **évaluation continue, év. permanente, év. en cours, év. endoformative, év. concurrente, év. concomitante, év. dynamique.**
- Différencier de : évaluation sommative

Evaluation interne (p. 38). Evaluation où l'évaluateur appartient à la structure responsable du projet [31].

- Terme voisin : **endo-évaluation.**
- Différencier de : autoévaluation, évaluation externe.

Evaluation négociée (p. 35). Evaluation dont le cahier des charges (*voir ce terme*) a fait l'objet d'une entente à l'interne (entre les acteurs) et à l'externe (bailleur de fonds, décideurs).

- Différencier de : évaluation participative.

Evaluation participative (p. 35). Elle accorde une place centrale à la participation de toutes les personnes concernées. Elle prend en considération leurs valeurs, leurs points de vue, leurs intérêts et leurs attentes, à toutes les étapes du processus d'évaluation [30].

- Terme voisin : **évaluation participante, év. pluraliste.**
- Différencier de : évaluation négociée.

Evaluation sommative (p. 21, 50). Elle détermine les résultats et l'impact du programme [9] et permet de comprendre les effets induits [18] et le programme dans leur globalité [14].

- Termes voisins : **évaluation-bilan, év. des effets, év. globale, év. récapitulative, év. comparative, év. analytique, év. intermédiaire, év. finale.**
- Différencier de : évaluation formative.

QUELQUES DEFINITIONS DE L'EVALUATION DE PROGRAMME
(voir aussi page précédente et p. 22)

- Une démarche pour *fournir aux intervenants et aux gestionnaires des renseignements qui les aideront à identifier les actions à poser afin de corriger certains aspects de leur programme* [4].
- *La comparaison d'un objet d'intérêt, d'un ensemble d'informations et d'une norme standard d'acceptabilité, d'un ensemble de critères adéquats, pour prendre une décision, juger entre diverses décisions possibles, sur base de rapport d'évaluation* [7].
- *Une démarche qui consiste à nourrir le corpus de connaissance des acteurs [...]. A partir de meilleures connaissances ils ajusteront leurs représentations, leurs comportements, leurs décisions.* [27].
- *Un processus d'analyse quantitative et/ou qualitative qui consiste à apprécier soit le déroulement d'une action ou d'un programme, soit à mesurer leurs effets (c'est-à-dire les effets spécifiques et les conséquences ou l'impact) ; c'est une démarche d'observation et de comparaison [...] entre les données recueillies et un référentiel* [1].
- *Une pratique qui s'efforce d'établir un lien explicite entre une dimension "cognitive" (acquérir de nouvelles connaissances), une dimension "normative" (jugement sur la réussite ou le bien fondé d'une action), une dimension "instrumentale" (produire du changement)* [14].
- *Apprécier le travail conduit, le visualiser et progresser dans les pratiques en intégrant les acquis et en analysant les obstacles et erreurs* [18].
- *Vérifier la pertinence et la cohérence du concept de départ, décrire la mise en oeuvre des moyens et de leur coût, évaluer la qualité de cette mise en oeuvre, mesurer le niveau d'atteinte des objectifs et en apprécier les effets et leur durabilité* [31].
- *Un processus de production d'un jugement de valeur pour conduire une action par des décisions. Une évaluation comporte donc trois phases ou trois composantes qui sont le processus, le produit [les conclusions et recommandations] et son usage. Ces trois composantes recouvrent trois enjeux de l'évaluation : connaître, juger et décider. La connaissance est l'enjeu de la science, qu'elle s'inspire de l'épidémiologie ou des sciences humaines. Le jugement, les valeurs, sont un enjeu de la philosophie, de la justice. La décision est l'enjeu de la politique* [16].
- *Porter un jugement de valeur sur une intervention en mettant en œuvre un dispositif permettant de fournir des informations scientifiquement valides et socialement légitimes sur une intervention ou sur n'importe laquelle de ses composantes de façon à ce que les différents acteurs concernés, qui peuvent avoir des champs de jugement différents, soient en mesure de prendre position sur l'intervention et de construire un jugement qui puisse se traduire en actions* [12].

Les encadrés ci-dessous donnent les définitions des types d'évaluation :

- Selon le moment du programme.
- Selon l'objet du programme.

EVALUATIONS SELON LE MOMENT DU PROGRAMME

Analyse de situation : voir ci-dessus page 88.

Evaluation en cours (p. 21). Evaluation qui a lieu pendant le programme. Voir aussi évaluation formative.

- Termes voisins : **évaluation concurrente, évaluation concomitante**.
- Différencier de : évaluation intermédiaire, évaluation finale, évaluation *ex-post*.

Evaluation intermédiaire (p. 21). Evaluation qui a lieu en fin de phase du programme. Voir aussi évaluation sommative.

- Différencier de : évaluation en cours, évaluation finale, évaluation *ex-post*.

Evaluation finale (p. 21). Evaluation qui a lieu à la fin du programme. Voir aussi évaluation sommative.

- Différencier de : évaluation en cours, évaluation intermédiaire, évaluation *ex-post*.

Evaluation ex-post (p. 21). Evaluation qui a lieu un certain temps après la fin du programme. Voir aussi évaluation sommative.

- Différencier de : évaluation en cours, évaluation intermédiaire, évaluation finale.

Evaluation post-décision (p. 17 et 21). Evaluation qui a lieu après la décision de lancer le programme, donc à n'importe quel stade de son avancement. Elle inclut les évaluations en cours, intermédiaire, finale et ex-post (*voir ces termes*).

- Termes voisins : **évaluation rétrospective, évaluation a posteriori**.
- Différencier de: analyse de situation.

EVALUATIONS SELON L'OBJET DU PROGRAMME

Evaluation de la cohérence (p. 48). Evaluation qui apprécie si le programme tel qu'il a été réalisé (ses ressources, ses activités et son implantation dans la population-cible) va bien dans le sens de ses objectifs de départ. Elle fait partie de l'évaluation de la mise en oeuvre (*voir ce terme*) : elle vérifie que les efforts déployés dans la réalisation du programme vont dans le sens des objectifs. *Voir aussi évaluation de la structure, évaluation du processus, évaluation d'implantation.*

- Termes voisins : **évaluation tactique, analyse du programme, analyse logique de l'intervention.**
- Différencier de : évaluation de l'écart.

Evaluation de l'écart (p. 47). Evaluation qui apprécie si le programme tel qu'il a été réalisé (ses ressources, ses activités et son implantation dans la population-cible) correspond à ce qui avait été prévu au départ. Elle fait partie de l'évaluation de la mise en oeuvre (*voir ce terme*) : elle vérifie que les efforts déployés dans la réalisation du programme sont conformes à la planification. *Voir aussi évaluation de la structure, du processus, d'implantation.* L'évaluation de l'atteinte des objectifs prévus peut aussi faire partie de l'évaluation de l'écart (*voir évaluation des résultats.*

- Termes voisins : **évaluation opérationnelle, évaluation normative, évaluation du déroulement de l'action.**
- Différencier de : évaluation de la cohérence.

Evaluation de la structure (p. 41). Evaluation de *l'adéquation des ressources [...] : la quantité, la qualité et l'agencement des ressources humaines, physiques et financières du programme* [35].

- Termes voisins : **évaluation des ressources, évaluation de l'input, des entrées, des intrants.**
- Différencier de : évaluation du processus, évaluation de l'implantation, évaluation des résultats, évaluation d'impact.

Evaluation des résultats (p. 42, 50). Elle *se rapporte aux objectifs du programme [...], c'est-à-dire [aux] changements liés aux problèmes visés par le programme et pour lesquels celui-ci a été élaboré* [9].

- Terme voisin : **évaluation de l'outcome, évaluation des objectifs.**
- Différencier de : évaluation du processus, évaluation du processus, évaluation de l'implantation, évaluation d'impact.

Continuez page suivante

EVALUATIONS SELON L'OBJET DU PROGRAMME (fin)

Evaluation des effets (p. 50). Un effet est *tout comportement ou tout événement dont on peut raisonnablement dire qu'il a été influencé par un aspect quelconque du programme* [7]. L'évaluation des effets s'intéresse au *degré d'atteinte des résultats escomptés et des autres conséquences* [35]. Elle inclut les évaluations des résultats, d'impact (sens restreint), de la pertinence (*voir ces termes*).

- Terme voisin : **évaluation d'impact** (sens large).
- Différencier de : évaluation des efforts.

Evaluation de la mise en oeuvre (p. 47). Ensemble des évaluations centrées sur la mise en œuvre du programme et l'atteinte des populations-cibles. Elle inclut les évaluations de la structure, du processus, d'implantation, de l'écart et de la cohérence (*voir ces termes*).

- Termes voisins : **évaluation des efforts, évaluation d'implantation** (sens large).
- Différencier de : évaluation des effets.

Evaluation de la pertinence (p. 51). Elle juge *de la contribution du programme à la solution des problèmes identifiés*. [35]. Elle fait partie de l'évaluation des effets.

- Terme voisin : **évaluation stratégique**.

Evaluation d'impact (p. 44, 50). Dans son sens restreint elle s'intéresse à *un grand nombre de changements directs et indirects, voulus ou non voulus, liés ou non aux objectifs explicites du programme, tant sur la population visée que sur la population entière* [9].

Dans son sens large : voir *évaluation des effets*.

- Différencier de : évaluation de la structure, évaluation du processus, évaluation de l'implantation, évaluation des résultats.

Evaluation d'implantation (p. 43, 47). Dans son sens restreint, elle conduit [...] à *l'identification de la population qui fait effectivement usage des services mis en place* [4] ou qui a été rejointe par les activités du programme. Elle peut être analysée de façon quantitative et qualitative.

Dans son sens large : voir *évaluation de la mise en œuvre*.

- Termes voisins : **évaluation de la couverture, évaluation de la pénétration**.
- Différencier de : évaluation de la structure, évaluation du processus, évaluation des résultats, évaluation d'impact.

Evaluation du processus (p. 42). Evaluation de *la justesse (appropriateness) des activités* [35]. Elle peut être analysée de façon quantitative et qualitative.

- Termes voisins : **évaluation des procédures, évaluation de l'output, des sorties, des extrants**.
- Différencier de : évaluation de la structure, évaluation de l'implantation, évaluation des résultats, évaluation d'impact.

Exposition (au programme) (p. 62). Participation d'une personne ou d'un groupe au programme. Par analogie avec la liaison entre une maladie et l'exposition à un facteur de risque, on recherche lors de l'évaluation d'un programme la liaison entre un effet et l'exposition au programme. *Voir aussi groupe intervention.*

Facteur de confusion (p. 62). *une variable qui est associée aussi bien à la variable indépendante (le programme) qu'à la variable dépendante (l'effet du programme).* [35]. Un facteur de confusion peut faire croire à tort qu'il existe une relation entre le programme et un effet, ou au contraire masquer une relation existante entre le programme et un effet.

- Terme voisin : **variable de confusion**.

Fiabilité (d'une mesure, d'un indicateur) (p. 24, 67). *Qui produit des résultats constants lorsqu'elle est appliquée de façon répétitive* [35] dans le temps ou par des personnes différentes.

Groupe intervention (p. 62). Population-cible du programme dans lequel on va mesurer les effets de celui-ci. Selon le type d'étude, ce peut être la population visée (*voir efficacité*), ou la population effectivement rejointe (*voir utilité*).

- Différencier de : groupe témoin.

Groupe témoin (p. 62). Groupe n'ayant pas bénéficié de l'intervention (ou selon le type d'étude, ayant bénéficié de l'intervention habituelle, ou avant intervention). Ce groupe est comparé au groupe intervention pour analyser l'effet propre du programme.

- Différencier de : groupe intervention.

Ici-ailleurs (étude) (p. 43). Catégorie de protocole qui compare deux groupes géographiquement distincts, un groupe intervention ("ici"), et un groupe témoin qui ne l'est pas ("ailleurs").

- Différencier de : avant-après (étude).

Impact (p. 44, 50). *Voir évaluation d'impact.*

Implicites (normes, critères) (p. 69). *Qui ne sont pas fixés à l'avance, mais plutôt laissés au jugement de l'évaluateur* [35].

- Différencier de : explicites (normes, critères)

Indicateur (p. 24, 66). *Variable qui aide à mesurer des changements* [34]. *Mesure dont les valeurs fournissent une indication incomplète mais utile sur un phénomène dont on cherche à apprécier l'évolution* [31].

- Différencier de : critère, norme.

Norme (p. 24, 68). *La valeur associée au critère [ou à l'indicateur] qu'on considère comme acceptable* [9].

- Terme voisin : **standard**.
- Différencier de : critère, indicateur.

Objectif (p. 42). *Énoncé formulé en termes clairs, précis et opérationnels qui décrit le résultat qu'on se propose d'atteindre. L'objectif doit préciser les personnes visées, la nature de la situation désirée [...] ainsi que la période de temps [...] nécessaire* [30]. On peut définir les objectifs d'un programme, mais aussi les objectifs de l'évaluation.

Évaluation des objectifs d'un programme : *voir évaluation du résultat.*

- Termes voisins : **but, résultat**.
- Différencier de : objet d'évaluation.

Objet d'évaluation (p. 41). *Ce sur quoi portera l'évaluation, c'est-à-dire la "chose" sur laquelle [on] souhaite porter un jugement* [30].

- Différencier de : objectif (d'évaluation).

Pertinence (p. 51). Voir évaluation de la pertinence.

Population-cible (p. 43). Groupe visé par le programme.

- Terme voisin : **groupe-cible, bénéficiaires**.

Programme (p. 14, 15). *L'ensemble des moyens (physiques, humains, financiers, symboliques) organisés dans un contexte spécifique, à un moment donné, pour produire des biens ou des services dans le but de modifier une situation problématique* [11]. Voir aussi évaluation de programme.

- Termes voisins : **projet, action, intervention**.
- Différencier de : technique, outil, pratique, politique.

Prospective (étude) (p.64). Etude où les groupes intervention et témoin (voir ces termes) sont constitués avant l'intervention et sont suivis durant celle-ci.

NE PAS CONFONDRE avec l'évaluation prospective ou analyse de situation (voir ce terme)

- Différencier de : rétrospective (étude).

Protocole d'évaluation (p. 63-65). Ensemble des règles définies avant une évaluation décrivant sa méthodologie ainsi que ses différentes étapes. Il est, dans ses grandes lignes, un élément du cahier des charges (voir ce terme). Il est plus détaillé à l'usage des personnes qui ont à réaliser l'évaluation. Dans le cas d'études épidémiologiques, on distingue, selon

le degré de contrôle de l'intervention, les protocoles expérimentaux, quasi-expérimentaux et pré-expérimentaux.

- Différencier de : cahier des charges.

Question d'évaluation (p. 21, 71, 72). *Interrogation se rapportant à l'objet d'évaluation et correspondant à ce que [on] veut réellement savoir à propos de l'objet retenu* [30].

Randomisation (p. 63). *Le tirage au sort [ou randomisation] veut dire ici que l'appartenance au groupe [10intervention ou témoin] est déterminé de façon aléatoire. Cette façon de procéder a pour but d'assurer que les deux groupes sont comparables pour tous les facteurs pronostiques (connus ou inconnus)*. [5].

- Termes voisins : **tirage au sort, tirage au hasard**.

Rétrospective (étude) (p.64). Etude se situant après l'intervention et s'intéressant à la façon dont les personnes ont été exposées à celle-ci. Voir aussi protocole pré-expérimental. **NE PAS CONFONDRE** avec évaluation rétrospective ou évaluation post-décision (voir ce terme)

- Différencier de : prospective (étude).

Suivi (p. 27). Collecte méthodique de données sur les activités et l'utilisation des ressources d'un programme.

- Termes voisins : **surveillance, monitoring, monitoring**.
- Différencier de : évaluation.

Supervision (p. 16). Processus de management des ressources humaines ayant pour but d'améliorer ou de maintenir la performance des personnes et des équipes.

- Différencier de : évaluation.

Tableau de bord (p. 27). Document regroupant une série d'indicateurs recueillis régulièrement pour le suivi d'un programme ou d'une population. Voir aussi suivi.

Tirage au sort (p. 63). Voir randomisation.

Triangulation (p. 68). Elle consiste à étudier la concordance ou la discordance entre différentes données, sources de données ou méthodes de collecte de données sur un même objet d'évaluation [21]. Voir aussi validité des mesures.

Utilité (Anglais : *efficacy*) (p. 43). Atteinte des objectifs dans la population rejointe par le programme (efficacité d'utilisation) ou dans des conditions d'intervention idéales (efficacité potentielle, efficacité théorique, efficacité d'essai).

- Termes voisins : **efficacité d'utilisation, efficacité potentielle, efficacité théorique, efficacité d'essai**.
- Différencier de : efficacité, efficience.

Validité des mesures (ou d'un indicateur) (p. 24, 66-67). Qualité d'une mesure à bien représenter l'objet dont elle doit rendre compte.

Validité interne (p. 62). En général, qualité de la méthodologie qui fait que les résultats de l'évaluation reflètent bien la réalité. En épidémiologie, qualité de la méthodologie et de l'analyse des données qui *permet de conclure que l'effet obtenu est réellement dû à la variable indépendante* [10ci, le programme évalué] *et non à des facteurs ou variables de confusion, l'effet de ces dernières ayant été pris en considération et contrôlé dans le* [protocole] [35]. Voir aussi protocole, facteur de confusion.

- Différencier de : validité externe.

Validité externe (p. 65). Qualité de la méthodologie de l'évaluation qui *permet de généraliser les résultats à une population plus grande ou à d'autres situations ou contextes plus larges*. [35].

- Terme voisin : **représentativité**.
- Différencier de : validité interne.

Viabilité (p. 45). *Possibilité de poursuivre, après la fin du programme, des activités développées grâce à ce projet. La viabilité s'apprécie en termes institutionnels, financiers et techniques* [31].

- Terme voisin : **pérennité**.

ANNEXE 2 : BIBLIOGRAPHIE

Les références soulignées sont les plus pertinentes et les plus accessibles pour approfondir la méthodologie de l'évaluation.

Toutes ces références sont consultables dans les centres de documentation de l'Espace régional de Santé Publique.

1. ANDEM- Service Formation. **Evaluation d'une action de santé publique : recommandations.** Paris : Agence Nationale pour le Développement de l'Evaluation Médicale (ANDEM), 1995, 48 p.
2. BANTUELLE M., MOREL J., DARGENT D. **L'évaluation, un outil au service du processus.** Bruxelles : asbl Santé, Communauté, Participation, 2001, 53 p.
3. BATTISTA R.N., TREMBLAY J. **Méthodes d'évaluation.** In : Dufresne J., Dumont F., Martin Y. (Eds). *Traité d'anthropologie médicale.* Québec : Presses de l'Université du Québec, 1985, p. 497-532.
4. BEAUDRY J. **L'évaluation de programme.** In : Begin C. (Ed). *Evaluation : un cadre conceptuel et une application au système de santé et de services sociaux.* Québec : Université Laval, 1986, p. 391-415.
5. BREART G., BOUYER J. **L'évaluation, l'exemple de la périnatalogie.** In : Brückner G. & Fassin D. (Eds). *Santé publique.* Paris : Ed. Marketing, 1989, p. 840-847.
6. BULART C., LAHAYE D., PADIEU R., RICAUX G. **L'évaluation en toxicomanie : questions préalables.** *Revue documentaire Toxibase*, 1995, 2, 1-19.
7. BURY J.A. **Education pour la santé.** Bruxelles : De Boeck – Wesmael, 1988, 224 p.
8. CFES. **La santé en action.** Paris : CFES, 1987, 112 p.
9. CHAMPAGNE F., CONTANDRIOPOULOS A.P., PINEAULT R. **Un cadre conceptuel pour l'évaluation des programmes de santé.** *Rev. Epidem. et Santé Publ.*, 1985, 33, 173-181.

10. CNAF (Caisse Nationale des Allocations Familiales). **L'action sociale est-elle efficace ?** Informations sociales, 1997, 57, 130 p.
11. CONTANDRIOPOULOS A.P., CHAMPAGNE F., DENIS J.L., PINEAULT R. **L'évaluation dans le domaine de la santé : concepts et méthodes.** In : Lebrun T., Saily J.C., Amouretti M. (Eds). *L'évaluation en matière de santé. Des concepts à la pratique. Journée d'étude avril 1991.* Lille : CRESGE, 1992, 480 p.
12. CONTANDRIOPOULOS A.P., CHAMPAGNE F., DENIS J.L., AVARGUES M.C. **L'évaluation dans le domaine de la santé : concepts et méthodes.** Rev. Epidem. et Santé Publ., 2000, 48, 517-539.
13. COPPE M., SCHOONBROODT C. **Guide pratique d'éducation pour la santé. Réflexion, expérimentation, et 50 fiches à l'usage des formateurs.** Bruxelles : De Boeck – Wesmael, 1992, 296 p.
14. CRDSU. **L'évaluation de la politique de la Ville.** Les Cahiers du CRDSU (Centre de Recherche sur le Développement Social Urbain), 1995, n°8, 44 p.
15. DECCACHE A. **Pour mieux choisir son évaluation... Définition et rôles des évaluations en éducation pour la santé.** Liège : APES (Association pour la Promotion de l'Education pour la Santé), 1989, 9 p.
16. DEMARTEAU M. **Les pratiques d'action en éducation pour la santé et leurs modes d'évaluation : réflexions critiques et questions pour la recherche sur l'évaluation.** Communication au colloque "L'évaluation en éducation pour la santé : entre épidémiologie et sciences humaines". Bordeaux, 15-17 septembre 1998, 14 p.
17. DEROSIERS H. & coll. **Cadre de pratique pour l'évaluation des programmes. Application en promotion de la santé et en toxicomanie.** Montréal : Gouvernement du Québec, 1998, 130 p. + fiches techniques.
18. DIREPS Grand Sud-Ouest. Commission méthodologique – Etudes – Recherches – **Evaluation. Outil d'aide à l'évaluation des programmes d'éducation et de promotion de la santé.** Albi : DIREPS Grand Sud-Ouest – Délégation Inter-Régionale pour l'Education et la Promotion de la Santé du Grand Sud-Ouest, Avril 1997, 99 p.

19. **FONDATION DE FRANCE. Charte de l'évaluation dans le cadre d'un travail de développement.** Paris : Fondation de France, 1995, 12 p.
20. **FONTAINE D. Evaluer les actions de santé dans le tiers-monde.** L'enfant en milieu tropical. 1992, n°198, 65 p.
21. **GAUDREAU L., LACELLE N. Manuel d'évaluation participative et négociée.** Montréal : Université du Québec - Service aux collectivités, 1999, 4 cahiers, 29+42+14+38 p.
22. **GERBAUD L. L'évaluation des actions de santé publique.** In : Pomey M.P., Poullier J.P., Lejeune B. (Eds). *Santé publique*. Paris : Ellipses, 2000, p. 201-208.
23. **GRAWITZ M. Méthodes des sciences sociales.** Paris : Dalloz, 2001, 1019 p.
24. **GREMY F., GOLDBERG M. Les sciences de l'information dans le domaine de la santé.** Paris : Faculté de médecine, 1981, 325 p.
25. **HARVOIS Y. Le contrôle, cet obscur objet du désir.** Pour, 1986, n°107 (L'évaluation au pouvoir), 116-119.
26. **INSERM (expertise collective). Education pour la santé des jeunes. Démarches et méthodes.** Paris, Editions Inserm, 2001, 247 p.
27. **INUDEL RHONE-ALPES. Evaluation dans le développement social urbain.** Lyon : Association régionale pour l'innovation urbaine et le développement local (Inudel), rencontre interrégionale, Février 1991, 146 p.
28. **KRÖGER C., WINTER H., SHAW R. Guide d'évaluation des actions de prévention des toxicomanies à l'intention des planificateurs et des évaluateurs de programme.** Munich : OEDT (Office européen des drogues et toxicomanies), 1998, 77 p.
29. **MANDERSCHIED J.C. Les évaluations en éducation pour la santé.** Thèse Université Montpellier 1, 1992, 344 p.
30. **MIDY F., VANIER C., GRANT M. Guide d'évaluation participative et de négociation.** Montréal : Université du Québec - Service aux collectivités, 1998, 4 fascicules : 18+32+26+26 p.

31. MINISTERE DE LA COOPERATION. **Guide des principes d'évaluation au Ministère de la Coopération, Pour évaluer plus et mieux.** Paris : Ministère de la Coopération, 1997, 85p.
32. MONTOYA-AGUILAR C. **Evaluation de l'efficacité des actions de santé publique.** Rev. Epidém. et Santé Publ., 1981, 29, 233-244.
33. NOCK F. **Petit guide de l'évaluation en promotion de la santé.** Paris : Mutualité française, 2000, 89 p.
34. ORGANISATION MONDIALE DE LA SANTE. **L'évaluation des programmes de santé. Principes directeurs.** Genève : O.M.S., 1981, 49 p.
35. PINEAULT R., DAVELUY C. **La planification de la santé. Concepts, méthodes, stratégies.** Montréal : Editions nouvelles, 1995, 480 p.
36. TARDY L. **Action de prévention du sida "Générer des dynamiques locales" 1995-1997. Evaluation des facteurs de pérennité.** Paris : Aides – Editions FNMF, 1997, 181 p.
37. WOMEN'S RESEARCH CENTRE. Maintenir le CAP. **Guide d'évaluation pour les groupes communautaires.** Vancouver : Women's Research Centre, 1990, 84 p. Traduction française : Montréal : Relais femmes (1265 rue Berri, #810, Montréal, Québec, H2L4X4)

Avec le soutien financier de :

