

1

Introduction générale à l'évaluation de programme

Valéry Ridde et Christian Dagenais

L'évaluation n'est pas chose nouvelle. On raconte dans l'Ancien Testament qu'à la cour du roi de Babylone, Nabuchodonosor, le chef des eunuques organisa une évaluation que l'on qualifiera ici de quasi-expérimentale. Il s'agissait d'étudier les effets différenciés d'un régime alimentaire sur la santé des enfants hébreux de la cour. Mais au-delà de la légende, il faut reconnaître que le champ de l'évaluation de programme s'est surtout développé dans les trente dernières années. L'origine de la pratique et de la recherche en la matière se situe essentiellement dans le domaine de l'éducation. Les enseignants se sont interrogés sur l'influence que pouvaient avoir leurs cours sur les connaissances acquises par leurs élèves. En changeant d'objet, passant des habiletés et des connaissances acquises par les individus aux programmes qui organisaient cette transmission, le champ de l'évaluation s'est transformé pour devenir beaucoup plus interdisciplinaire. Ainsi, quand on cherche à appréhender pour la première fois le domaine de l'évaluation de programme, on s'aperçoit qu'il existe de multiples façons de le concevoir. Si vous consultez cinq ouvrages de référence, vous risquez d'y trouver cinq manières distinctes de définir l'évaluation et ses concepts. Il s'agit en effet d'une discipline dont les concepts et la pratique n'ont pas encore été clairement codifiés. Voilà

pourquoi plusieurs sociétés nationales d'évaluation cherchent encore à se doter de chartes, de codes de déontologie, de normes de pratique ou de compétences essentielles, alors que d'autres en ont adopté récemment. Voilà aussi pourquoi, lorsque l'on souhaite renforcer les capacités et les pratiques évaluatives au sein d'une organisation, il est toujours bon de se doter d'une politique commune d'évaluation, voire d'un glossaire des termes que l'on utilisera (le présent ouvrage en propose un en page 327).

La présente introduction est basée sur notre propre expérience de l'enseignement et sur notre pratique de l'évaluation dans un grand nombre de contextes et sur diverses problématiques. Il ne s'agit évidemment pas de suggérer une nomenclature définitive ou supérieure aux autres. Notre objectif est bien plus modeste et consiste à présenter au lecteur le champ de l'évaluation de programme afin de faciliter sa lecture des différents chapitres de cet ouvrage.

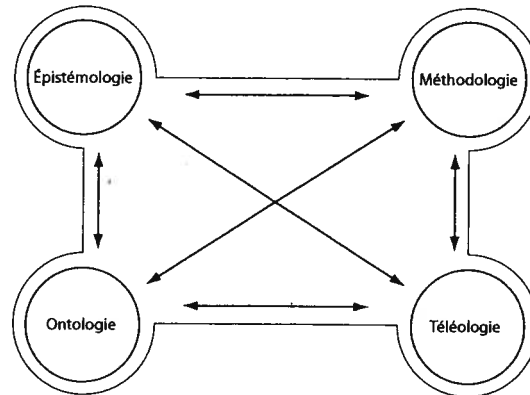
Les courants de pensée : une brève histoire de l'évaluation

Les courants de pensée en évaluation sont nombreux, et la façon dont les évaluateurs se représentent le monde (leur « paradigme ») conditionne largement les questions qu'ils se posent et les stratégies qu'ils emploient pour y répondre (Bégin *et al.*, 1999). Nos valeurs, croyances et intérêts interfèrent assurément dans nos pratiques professionnelles. Dans le cas qui nous occupe, ils déterminent les choix théoriques et méthodologiques de la pratique de l'évaluation. Il ne s'agit pas de discuter longuement des grands paradigmes épistémologiques, mais il est essentiel d'en dire quelques mots pour que le lecteur, évaluateur actuel ou en devenir, soit en mesure de s'interroger à son tour sur ses propres choix¹. Pour le définir simplement, un paradigme, selon la définition classique de Thomas Kuhn, est un système de croyances partagé par une majorité d'individus œuvrant au sein d'une même discipline. Ce système de référence est un cadre normatif qui oriente les questions, les méthodes et la manière d'attribuer un sens aux données que l'on recueille. À la suite de Levy, Gendron (2001) propose qu'un paradigme soit constitué de quatre champs interreliés.

1. Pour approfondir cette réflexion, nous suggérons la lecture de l'excellent manuel d'introduction à l'épistémologie de Fourez (2003).

FIGURE 1.1

Les quatre champs constitutifs d'un paradigme



Source : Gendron, 2001.

- Le champ épistémologique renvoie à la vision du monde de l'évaluateur, la légitimité des analyses qu'il produit et surtout sa relation (et sa distance) avec les données qu'il recueille et analyse (subjectivité *par opposition* à objectivité).
- Le champ ontologique englobe la nature du monde vécu et la compréhension qu'on a de la manipulation possible par l'évaluateur des entités physiques ou mentales. C'est de la construction de la réalité dont il est ici question (une seule réalité *par opposition* à plusieurs).
- Le champ méthodologique rassemble les techniques que l'évaluateur emploie pour appréhender le monde vécu.
- Le champ téléologique regroupe les questions relatives aux finalités des évaluations ainsi qu'aux intérêts et intentions des évaluateurs d'un programme.

Le paradigme au sein duquel l'évaluateur s'inscrit conditionne les méthodes utilisées, par exemple, pour répondre aux questions d'évaluation posées par les responsables d'un programme. Certains évaluateurs, dits positivistes, ne pourront croire à l'efficacité d'un programme que dans la mesure où il aura été possible d'en comparer les effets, comme dans un laboratoire, entre un groupe de participants (cas) et un groupe de non-participants (témoins), l'appartenance de ces personnes à ces groupes ayant été déterminée aléatoirement (voir le chapitre 7). D'autres évaluateurs,

plutôt constructivistes, affirmeront à l'inverse qu'il est impossible d'isoler les effets d'un programme de son environnement, et qu'il faut employer une méthodologie permettant de relever ces interactions. Ils considèrent ainsi que dans le cas des programmes sociaux, il est impossible de reproduire des conditions de laboratoire. Ils soutiennent aussi que toute connaissance sur les effets d'un programme est construite conjointement par les acteurs sociaux concernés. Ces évaluateurs constructivistes assument pleinement la subjectivité de leur jugement, puisqu'ils estiment que l'objectivité dont se réclame le camp positiviste n'est pas atteignable.

Les différents courants de pensée de l'évaluation se sont développés tout au long de son histoire. L'une des manières de décrire cette histoire, car nous sommes d'accord avec Shaw *et ses collègues* (2006) pour dire qu'il en existe plusieurs, est de suggérer l'existence, parfois concomitante, de quatre générations d'évaluation (Guba et Lincoln, 1987). La première génération est celle de la mesure, dont l'évaluateur est un spécialiste. La deuxième est plutôt de l'ordre de la description. L'évaluateur s'attache surtout aux processus qui permettent la survenue des effets constatés. Dans la troisième génération, le défi de l'évaluateur est de porter un jugement sur les résultats obtenus en regard des objectifs initialement fixés. Enfin, dans les années 1970 est apparue une quatrième génération, qui se veut bien plus interactive. Les évaluateurs usent de la négociation, deviennent des facilitateurs entre les instances impliquées. Il s'agit de reconnaître non seulement le point de vue et les valeurs de l'ensemble des acteurs concernés, mais aussi de faire en sorte qu'ils fassent partie intégrante de ce processus. Autrement dit, il s'agit d'évaluations pluralistes dont l'approche participative, décrite plus bas, est l'une des stratégies possibles.

L'évaluation de programme : une définition

Plutôt que de commencer par une définition universitaire, nous vous proposons un bref séjour au soleil, dans un pays de l'Afrique de l'Ouest. Par une radieuse journée de janvier, un expert québécois anime une séance de formation à l'évaluation de programme au Burkina Faso. Cette séance est organisée à l'intention des sages-femmes et des intervenantes d'une organisation non gouvernementale (ONG) qui met en œuvre un programme de maternité dite « à moindre risque ». En commençant sa formation sur l'évaluation de programme, l'expert interroge les partici-

pantes sur les mots qu'elles associent au terme « évaluation ». On évoque le contrôle, l'audit, la mesure, la vérification, etc. Des Canadiens présents dans la salle parlent de l'idée de reddition de comptes. La teneur de ce lexique ne surprend guère le formateur, puisque aujourd'hui encore, beaucoup associent avant tout l'évaluation à des fonctions de contrôle qui attribuent de bonnes ou de mauvaises notes à un programme. Puis, l'expert demande : « Quelle température fait-il aujourd'hui ? » Les agentes de la santé trouvent qu'il fait plutôt froid, les expatriées canadiennes de l'ONG, qu'il fait chaud, et l'expert québécois qui vient d'arriver juge qu'il fait très, très chaud. « Comment avez-vous fait pour juger de la température et dire qu'il fait froid ou chaud ? » Chacun répond de manière différente à la question de l'expert : « c'est l'hiver », « j'ai regardé le thermomètre », « j'ai comparé avec la température de la veille », « je transpire beaucoup plus aujourd'hui qu'hier ». Puis, les participantes admettent toutes avoir choisi leurs vêtements en fonction de la température de ce matin-là. Les personnes présentes à cet atelier de formation ont alors pris conscience du fait que l'évaluation est un acte permettant de porter un jugement (froid, chaud, tempéré) selon une démarche critique fondée sur une collecte systématique de données (température, sudation, saison) à propos de multiples objets (la température) dans le but de prendre des décisions (comment s'habiller).

Certains auteurs soutiennent que le jugement évaluatif à l'égard d'un programme doit être posé selon une démarche scientifique ou en ayant recours à des méthodes scientifiques. Nous préférons ne pas employer le terme « scientifique », car dans notre pratique de l'évaluation, le mot a toujours eu tendance à créer une certaine distance avec les parties impliquées dans l'évaluation. L'évaluateur est alors rapidement perçu comme un « chercheur », avec toutes les connotations qui, à tort ou à raison, s'y rattachent. En outre, évoquer la science fait perdurer la difficulté qu'ont la plupart des auteurs à différencier la recherche de l'évaluation (chapitre 2). La différence fondamentale, selon nous, entre la recherche et l'évaluation est que la finalité principale de la première est de produire des connaissances, alors que le but de la seconde est de fournir une information utile sur le fonctionnement d'un programme ou pour permettre une prise de décision le concernant.

Évidemment, les choses ne sont pas toujours aussi tranchées : on peut parfois produire des connaissances à l'aide d'une évaluation, ou prendre

des décisions à la suite de travaux de recherche. Mais il reste que le choix de ne pas la définir explicitement comme une science ne signifie pas que l'évaluation soit une démarche frivole où le jugement que l'on porte sur l'objet programme n'est pas fondé. Au contraire, la rigueur et la systématisation en sont les deux valeurs-phares.

L'analogie climatique est simple, mais suffisamment claire pour définir l'acte d'évaluation et s'assurer que les personnes avec qui nous allons travailler ne la perçoivent pas comme un processus de contrôle (bien que cela puisse parfois être le cas). Évaluer la température est un peu plus simple que d'évaluer un programme. Mais comment peut-on définir un programme ? Ici encore, une multitude de définitions se croisent. Nous retenons celle de Plante (1994). Bien qu'un peu longue, elle a le mérite de préciser clairement les contours de ce qui constitue un programme.

[Un programme est un] ensemble cohérent, organisé et structuré d'objectifs, de moyens et de personnes qui l'animent. Il se justifie sur la base de besoins définis comme une carence ou un manque qui affecte des individus, une collectivité ou une société. Il est sous le contrôle d'une ou de plusieurs personnes responsables de la qualité de sa formulation et de son fonctionnement. Il est mis en place pour transformer des choses ou l'état d'une chose.

Le maître mot de cette définition est transformation. Un programme est donc un ensemble organisé et cohérent d'activités dont l'objectif n'est pas d'exister pour lui-même, mais plutôt de produire des changements pour les participants potentiels. En effet, il ne sert à rien, ou alors à d'autres choses que celles qui sont évoquées ici, de développer un programme de maternité à moindre risque si celui-ci vise seulement à construire des maternités, à former des sages-femmes ou à rédiger des protocoles thérapeutiques. Ce que l'on vise avant tout, c'est l'obtention à court terme d'une meilleure qualité des soins pour les parturientes et leurs nouveau-nés, et à plus long terme, une réduction de la mortalité maternelle et infantile. Il faut aussi comprendre que tout programme peut produire des changements inattendus dont les évaluateurs devront aussi rendre compte. Ces effets peuvent avoir été anticipés ou non, et être bénéfiques, neutres ou nuisibles pour les participants.

Même si les frontières ne sont pas toujours faciles à tracer, cet ouvrage traite de l'évaluation de programme, et laisse donc de côté l'évaluation des pratiques professionnelles, des connaissances individuelles ou encore des politiques publiques.

Au-delà des aspects conceptuels de ces définitions de ce qu'est un programme, il faut noter combien les acteurs sociaux jouent un rôle crucial dans sa dynamique. C'est par l'exercice du pouvoir que se réalisent ou non les programmes, bien loin de se déployer d'eux-mêmes (Crozier et Friedberg, 1977; Lemieux, 2002). Il faut donc être très attentif au jeu des acteurs, aux relations de pouvoir et aux interactions sociales. Aussi plusieurs acteurs sont-ils impliqués lorsqu'un programme est défini, puis mis en œuvre, mais également lorsqu'il est évalué. L'évaluation, en tant qu'activité humaine, est source de tensions et de relations de pouvoir. Il ne faudrait donc pas voir un programme et son évaluation comme une activité mécanique, uniquement technique, où ne nous intéressent que les objectifs et les indicateurs.

Si un programme peut aussi être conçu intellectuellement comme la résultante d'un ordre négocié entre des acteurs sociaux, selon l'expression d'Anselm Strauss, pour les fins de cette introduction, il nous faut avoir recours à des subterfuges didactiques pour expliciter les concepts théoriques. Les différents termes que nous suggérons sont donc essentiellement heuristiques, car ils expliquent la matière enseignée. Il ne s'agit cependant pas de nier la complexité des programmes, dirait Edgar Morin, et des interactions sociales. Les tenants de l'évaluation réaliste perçoivent le changement social comme un système de transformation, un système ouvert qui est le produit d'une myriade d'éléments et de forces impossibles à quantifier (Pawson, 2006; Ridde et Guichard, 2008; Blaise *et al.*, 2009). On aura donc compris que la vision du monde et de la science qu'ont les évaluateurs – leur perspective paradigmatique – conditionne leur manière de percevoir et d'évaluer un programme.

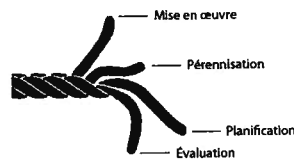
L'évaluation n'est pas l'étape finale d'un programme

La vision des programmes qui veut que ceux-ci doivent être adaptés vient contredire celle qui décrit leur vie selon un cycle inéluctable allant de l'identification de besoins à l'évaluation, en passant par les étapes de la planification et de la mise en œuvre. La plupart des manuels de gestion de programme illustrent encore le processus des programmes par un cercle représentant ces différentes étapes jointes par des flèches. Les fondements de cette conception remontent assurément aux «étapistes» des années 1950 qui étudiaient les politiques publiques. L'évaluation y est souvent l'étape ultime du processus. D'autres ajoutent parfois la pérennité ou encore la capitalisation pour clore la boucle.

Nous estimons pour notre part que le processus des programmes n'est ni linéaire ni cyclique. À la suite des propositions théoriques énoncées par Pierre Pluye et ses collaborateurs (2004), il faudrait concevoir un programme comme étant composé d'une série de quatre sous-processus concomitants et interdépendants que sont la planification, la mise en œuvre, l'évaluation et la pérennisation. Ces quatre sous-processus pourraient ainsi être représentés par les quatre fils d'une même corde programmatique.

FIGURE 1.2

Concomitance et interdépendance des sous-processus d'un programme



Adapté de Pluye *et al.*, 2002.

La planification permet de prévoir l'allocation des ressources nécessaires pour atteindre des objectifs précis grâce à la mise en œuvre d'activités préalablement définies. Quant à l'évaluation d'un programme, adaptant des définitions plus classiques et formelles (Grembowski, 2001; Mathison, 2004; Patton, 1997) que celle que nous avons évoquée, nous dirons qu'il s'agit d'un processus systématique de collecte d'informations utilisables pour porter un jugement sur un programme et ses composantes, dans le but de prendre des décisions. Enfin, alors que la mise en œuvre d'un programme consiste en la mobilisation de ressources pour organiser des activités afin d'atteindre des objectifs, la pérennisation concerne le devenir des programmes. La pérennisation est en effet centrée sur le processus qui permet la continuation des activités et des effets des programmes. La pérennité, elle, est le résultat du processus de pérennisation – résultat qui se manifeste dans les routines organisationnelles – et peut être évaluée de manière transversale, à tout moment et ponctuellement, en fonction de degrés de pérennité. Fait méconnu : les actions favorables à la pérennité doivent démarrer dès la mise en œuvre du programme, et non à la fin de celui-ci (Pluye *et al.*, 2005). Nous n'irons pas plus loin dans cette réflexion sur la pérennisation et la pérennité, certes

nouvelle dans le domaine, mais dont les détails conceptuels et les illustrations empiriques sont présentés ailleurs (Pluye *et al.*, 2004; Ridde, Pluye et Queuille, 2006).

Retenons que ces quatre sous-processus sont concomitants. Ceux qui ont été impliqués dans des programmes savent bien que lorsque l'on met en place un programme (mise en œuvre), on s'interroge toujours sur ce qui se passe (évaluation). On revoit toujours ce qu'il a été prévu de faire (planification) en se préoccupant, ce qui demeure cependant souvent le maillon faible des programmes, de la façon de procéder pour assurer l'avenir de ces derniers (pérennisation).

La finalité d'une évaluation

Dans un livre publié au milieu des années 1980, Patton établit ce qu'il a nommé une « soupe à l'alphabet de l'évaluation » (*alphabet soup*), au sein de laquelle il finit par trouver 100 types d'évaluation de programme. Plus récemment, il a réduit ce nombre à 58. Stufflebeam (2001) suggère quant à lui de regrouper 22 types d'évaluation en quatre grandes catégories :

- pseudo-évaluations : relations publiques ;
- quasi-évaluations : orientées vers les questions et les méthodes ;
- amélioration et imputabilité : orientées vers le mérite et la valeur ;
- plaidoyer social : orienté vers la différence dans la société.

Malgré ces distinctions variées, il est possible de suggérer des différences selon les finalités des évaluations. De manière classique, on estime que les évaluations poursuivent deux grands objectifs : rendre des comptes ou améliorer un programme. Le premier type d'objectifs permet aux responsables d'un programme de rendre des comptes dans une perspective d'imputabilité. On parle parfois d'une évaluation sommative, car elle est bien souvent réalisée à la fin d'un programme, soit *ex post*. Le second type d'objectifs produit des informations utiles pour corriger la manière dont les activités d'un programme sont mises en œuvre. On parle aussi d'évaluation formative, car celle-ci est habituellement effectuée pendant la mise en œuvre d'un programme. Au-delà de cette grande distinction dans la finalité d'une évaluation, quelques auteurs ont désiré en ajouter deux autres. D'un côté, l'évaluation serait aussi utile pour améliorer l'état des connaissances à propos de l'efficacité d'une intervention, pour

construire une théorie et pour publier des articles scientifiques (Patton, 1997). Elle se rapproche ainsi de la pratique de recherche. D'un autre côté, la finalité de l'évaluation de programme ne serait pas uniquement liée aux programmes et à leurs composantes, mais aussi à l'amélioration des conditions sociales et collectives, ce que disait déjà Carol Weiss au début des années 1970. L'évaluation devient alors un moyen, elle n'est pas une fin en soi. Certains cherchent même à renforcer le pouvoir d'agir (*empowerment*) des participants au programme et à l'évaluation, ce dont nous traiterons au chapitre 14.

Planifier une évaluation

Si l'on fait abstraction de la concomitance des sous-processus, un processus évaluatif se déroule habituellement en deux phases, parfois successives mais qui peuvent être aussi parallèles ou en boucle. Il s'agit de la phase préévaluative et de la phase évaluative. Certains auteurs leur ajoutent une troisième phase, celle de l'utilisation des résultats des évaluations (chapitre 10) pour la prise de décisions (Grembowski, 2001). Notons cependant que tous les évaluateurs ne semblent pas être d'accord pour endosser cette responsabilité.

Le nom de la première phase est une traduction du concept anglophone d'*evaluability assessment*. L'étape préévaluative (Beaudry et Gauthier, 1992) permet de s'assurer, non seulement qu'il est possible de réaliser l'évaluation d'un programme, mais aussi que les évaluateurs disposent de toutes les informations sur la demande d'évaluation. Wholey propose quant à lui deux idées à cet égard. Il distingue l'étude de la faisabilité de l'évaluation (la présence des conditions permettant de réaliser l'évaluation), d'une part, et l'analyse du caractère évaluable d'un programme (essentiellement liée à la cohérence du modèle logique), d'autre part. Ainsi, les enjeux de l'évaluation, qui sont parfois politiques (chapitre 4), doivent être appréhendés, tout comme doivent être anticipés les conflits potentiels. Les tenants d'une évaluation centrée sur l'utilisation des résultats, comme Patton (1997) (chapitre 8), et ceux d'une évaluation réactive, comme Stake (2004), s'interrogeront immédiatement sur ce qui préoccupe les parties prenantes et sur la manière dont elles entendent mettre les recommandations en œuvre.

Ensuite, le cœur de cette démarche préévaluative est de vérifier que les individus concernés par l'évaluation savent précisément en quoi consiste

le programme qu'il s'agit d'évaluer. Si nombre d'organisations mettent des interventions en œuvre et souhaitent qu'un jugement soit posé sur ces dernières, il est rare qu'elles disposent d'une description détaillée du programme, de ses composantes et de ses objectifs. Notre expérience montre que cette lacune constitue la norme plus que l'exception. Aussi le travail fondamental à réaliser d'emblée est-il de reconstituer la logique du programme (chapitre 6) tel qu'il a effectivement été mis en œuvre. Même si cette logique est parfois présente sous diverses formes dans les documents de planification, il existe toujours une différence entre ce que l'on a prévu de faire et ce que l'on fait vraiment ; c'est ce que mesure l'évaluation d'implantation (chapitre 20). Il demeure donc indispensable de décrire de manière consensuelle le programme dont il est question dans l'évaluation ainsi que ses composantes. Si l'on ne sait pas exactement quels sont les effets attendus d'un programme, comment en évaluer l'efficacité ? Si l'on ne connaît pas précisément l'enchaînement des activités et les éventuels liens de causalité avec des effets attendus, il est impossible d'attribuer au programme la responsabilité des changements observés.

Selon les écoles de pensée et leur terminologie, il faudra donc disposer, à la fin de cette préparation, d'un modèle logique, d'une théorie ou de la logique du programme à évaluer. Certains auteurs établissent une différence entre le modèle logique opérationnel d'un programme et son modèle théorique conceptuel. Dans la même veine, Chen (2005) différencie le modèle de changements (hypothèses descriptives à l'égard des relations causales qui expliquent l'atteinte des objectifs) du modèle d'action (hypothèses prescriptives à l'égard des composantes du programme qui vont faire en sorte qu'il fonctionne). Mais retenons surtout ici que la description précise d'un programme et des liens entre ses composantes est indispensable à tout processus évaluatif.

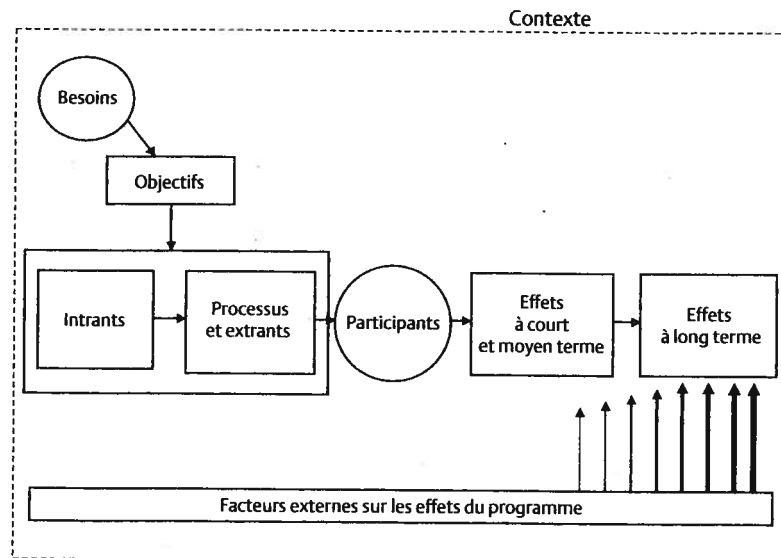
Pour y parvenir, il est souvent utile de faire une représentation graphique des relations entre les activités prévues et les effets attendus, dont on choisira la forme et les symboles. Les chapitres de la seconde partie de cet ouvrage fournissent des exemples intéressants à ce sujet. L'important, dans ces images, est que les relations soient claires et suffisamment précises pour orienter la manière dont les questions d'évaluation seront posées.

Ces représentations peuvent également être utiles pour montrer que la logique d'un programme ne produira pas d'effets tant que, par exemple,

la manière dont celui-ci est conçu ira à l'encontre des connaissances dont nous disposons sur les manières d'intervenir dans un domaine particulier. Par exemple, les données probantes actuelles en matière de réduction de la mortalité maternelle nous apprennent que la formation d'accoucheuses traditionnelles et la tenue de consultations prénatales sont relativement inefficaces. Si la construction de la logique d'un programme par ses acteurs met en évidence qu'ils développent leur programme en suivant ces liens de causalité, l'évaluateur pourra d'emblée porter un jugement sur cette «logique illogique». Dans un tel cas de figure, la phase préévaluative permet d'éviter la mise en œuvre d'une évaluation des effets et de suggérer, par exemple, de réaliser une évaluation des possibilités d'action afin de revoir la logique du programme. Nous proposons en figure 1.3 une représentation graphique des éléments composant un programme, mais il en existe beaucoup d'autres.

FIGURE 1.3

Représentation générique de la logique d'une intervention



Si l'on conserve l'exemple du programme de maternité à moindre risque, les besoins (mortalité, morbidité) des participantes (les femmes en âge de procréer) impliquent la mise en œuvre d'un programme qui

visé des objectifs (réduire la mortalité maternelle) dans un contexte particulier (celui d'une région de l'Afrique). Ce programme est composé d'*intrants* (les ressources financières et humaines nécessaires à la mise en place des activités) et de *processus* (formation, chantier de construction) qui produisent des *extrants* (livres, bâtiments de la maternité) dans le but de générer des effets à court terme (qualité des soins) sur les participantes (parturientes et sages-femmes) ou des effets à plus long terme (réduction du taux de mortalité maternelle).

Dans la figure 1.3, les participants sont visuellement bien séparés des processus afin de souligner le fait qu'un programme se justifie sur la base des changements qu'il cherche à induire en faveur de ses bénéficiaires. Les flèches horizontales mettent l'accent sur la chaîne des résultats, les liens de causes à effets. L'utilisation de flèches de plus en plus grandes au-dessous des effets permet d'insister sur l'importance de rester modeste quant à l'atteinte des effets. Car plus on se dirige vers la droite de la figure (vers les effets à long terme), plus des facteurs externes (autres programmes, acteurs, changements contextuels) sont susceptibles d'être à l'origine de ces effets, et donc, moins on peut en attribuer le bénéfice au programme. Ce n'est évidemment pas un programme de santé publique, quelle que soit son ampleur, qui pourra contribuer seul à réduire la mortalité maternelle. Par exemple, on sait parfaitement que l'accroissement du taux de scolarisation des petites filles aura sur la mortalité maternelle un effet à long terme aussi important, sinon plus, qu'un programme de santé publique.

Une fois la logique de l'intervention circonscrite, il faut pouvoir déterminer les questions d'évaluation auxquelles il s'agira de répondre. Dans de nombreux contextes, les commanditaires d'une évaluation veulent tout savoir sur tout, souvent en peu de temps et avec un budget restreint. Il est donc souvent indispensable de sérier les demandes et de vérifier la faisabilité de l'évaluation. On pourra mettre à profit l'approche suggérée au chapitre 9 pour mieux cerner les questions d'évaluation à l'aide des critères de temps, de ressources et d'accès aux données, ainsi que des critères politiques ou stratégiques.

Nous avons mis cette grille en pratique en 2005, lors d'une évaluation réalisée à Haïti. Une fois bien compris les enjeux et les besoins, une liste de la totalité des questions d'évaluation qui permettraient de répondre à l'ensemble des requêtes formulées dans le mandat que nous avons reçu

a été dressée. Devant l'ampleur de la tâche et l'impossibilité manifeste de répondre à toutes ces interrogations, une liste des questions prioritaires a été constituée. Deux séries de critères ont été successivement appliquées pour effectuer cette sélection. La première série était d'ordre technique, reposant sur trois indicateurs : le temps, le budget et l'accès aux données. Certaines questions ont aussi été immédiatement écartées pour des raisons méthodologiques. La seconde série de critères était d'ordre stratégique et liée à l'utilisation potentielle des résultats de cette évaluation.

Une fois réglées ces questions cruciales, il faut élaborer un devis d'évaluation. Des exemples sont fournis dans la seconde partie de cet ouvrage. Encore une fois, afin de ne pas créer de distance inutile entre les évaluateurs et les parties prenantes, il est parfois préférable d'employer le terme « plan » plutôt que « protocole » ou « devis » d'évaluation, qui connotent le monde de la recherche. Pour chacune des questions d'évaluation retenues, il faut déterminer :

- les données nécessaires pour y répondre ;
- les instruments de collecte de ces données ;
- les personnes qui détiennent ces données et les lieux où elles se trouvent ;
- le moment où les instruments de collecte seront utilisés et par qui ;
- les méthodes d'analyse des données.

Ensuite, une évaluation de programme se gère comme toute action : il faut prévoir une liste des tâches, un calendrier d'exécution, une description des ressources humaines et un budget. Nous avons dit qu'une évaluation était surtout utile pour la prise de décisions ; par conséquent, il est indispensable, dès la planification, de s'interroger sur la manière dont seront reçus les résultats (chapitre 20). Il faut réfléchir aux stratégies que l'on entendra mettre en œuvre pour communiquer ces résultats et faire en sorte que les personnes concernées puissent se les approprier. Une évaluation n'est réellement utile que si ses résultats sont utilisables et utilisés.

La mise en œuvre d'une évaluation : types et approches

Nous avons choisi de négliger ici le côté théorique de la mise en œuvre d'une évaluation. En effet, plutôt que de discourir sur la réalisation, nous

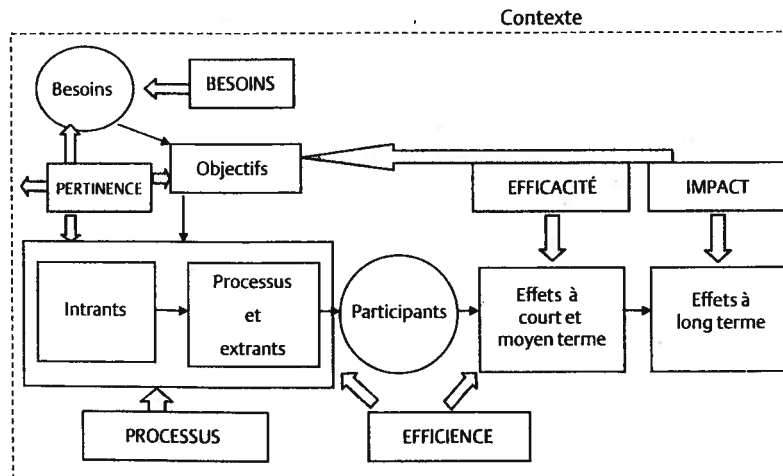
avons préféré présenter au lecteur des études de cas concrètes offrant des réponses à une question que les étudiants posent souvent : « *Comment fait-on l'évaluation d'un programme ?* » Le lecteur trouvera donc dans la seconde partie de cet ouvrage plusieurs illustrations de démarches évaluatives. Pour accroître l'étendue de ces exemples, nous présenterons des évaluations effectuées dans plusieurs domaines comme l'action sociale, l'éducation, la santé ou l'administration, évaluations de différents types et qui s'appuient sur des approches distinctes.

Les types d'évaluation

La représentation schématique de la figure 1.4 est une introduction efficace aux différents types d'évaluation ; le type d'évaluation étant ici défini par l'objet sur lequel porte le jugement de l'évaluateur. Dans le cas de l'évaluation d'un programme, cet objet peut être une composante du programme ou encore la relation entre ses composantes. Il serait trop long ici de passer en revue les différents types d'évaluation, aussi en donnerons-nous simplement quelques exemples.

FIGURE 1.4

Les types d'évaluation et leurs relations avec les composantes d'un programme



Le tableau 1.1 propose quelques exemples de questions choisis en fonction des divers types d'évaluation. Les études de cas de la seconde partie de cet ouvrage fournissent plusieurs illustrations empiriques de ces questions génériques.

TABLEAU 1.1

Types d'évaluation et questions génériques

Types d'évaluation	Exemples de questions génériques
Évaluation des besoins	Quels besoins le programme vise-t-il à satisfaire? Les besoins des participants ont-ils évolué depuis le début de la mise en œuvre? Voir chapitre 12.
Évaluation de la pertinence	Le programme répond-il aux besoins des participants? Le programme prend-il en compte le contexte de sa mise en œuvre?
Évaluation du processus	Le programme est-il mis en œuvre tel qu'il a été planifié? Quels facteurs facilitent ou entravent la mise en œuvre du programme? Voir chapitres 13 et 15.
Évaluation de l'efficacité	Les effets constatés à court terme correspondent-ils aux objectifs initialement fixés? Comment les effets relevés chez les participants au programme se comparent-ils avec la situation des non-participants? Voir chapitres 14, 15, 16 et 17.
Évaluation de l'impact	Les effets constatés à long terme correspondent-ils aux objectifs initialement fixés?
Évaluation de l'efficacité	Quel est le rapport entre les ressources et les effets du programme? Ce programme est-il plus onéreux qu'un autre qui produirait le même effet?

Notons que de nombreuses autres typologies sont proposées dans la littérature sur l'évaluation. Mais peu importe celle qui est privilégiée, il est toujours essentiel de bien analyser le contexte: selon sa nature, un programme n'aura pas les mêmes effets et ne sera pas mis en œuvre de la même manière. Il faut aussi rappeler que des effets non prévus, qu'ils soient favorables ou non aux participants, peuvent être générés, et ils doivent impérativement être relevés.

Chen (2005) estime que certains types d'évaluation s'appliquent mieux à certains moments de la vie d'un programme. Mais d'autres types d'évaluation peuvent aussi être employés. Par exemple, lorsque l'on souhaite concentrer les ressources sur des activités qui ont plus de chances de réussir, il est parfois utile de réaliser une évaluation des *possibilités d'action*. L'évaluation de la *cohérence* des activités d'un programme est aussi quelquefois requise. Certains distinguent l'évaluation de l'*implantation* (décrire ce qui se passe) de l'évaluation des processus (comprendre la

dynamique interne). Au sein de la catégorie de l'évaluation d'efficacité, aussi nommée « évaluation économique », il est utile de distinguer quatre types particuliers d'évaluation : minimisation des coûts, coût-efficacité, coût-utilité et coût-bénéfices (Drummond *et al.*, 1998). Il ne faut pas oublier, enfin, la possibilité de réaliser une évaluation de l'équité (Potvin *et al.*, 2008 ; Ridde *et al.*, 2009), ce qui est encore trop rarement fait.

Les approches évaluatives

Pour répondre aux questions, l'évaluateur peut avoir recours à plusieurs approches évaluatives. Le terme d'approche renvoie ici à la manière dont les experts interagissent avec les individus concernés par l'évaluation. On peut aussi définir l'approche comme étant la façon dont l'évaluateur assume son rôle. L'approche utilisée varie en fonction des situations, de l'organisation dans laquelle l'évaluateur évolue, de ses caractéristiques personnelles et des besoins précisés par les personnes concernées. Il n'y a pas d'approche qui soit *a priori* meilleure qu'une autre. Il s'agit de choisir celle qui paraît la plus pertinente dans le cas étudié. En forçant un peu le trait de l'idéal-type wébérien, il est possible de mettre en évidence les éléments fondamentaux qui caractérisent trois approches particulières.

L'évaluateur qui emploie une *approche directive* a un rôle d'expert et reste relativement neutre et distant par rapport à l'objet évalué. Il dirige l'ensemble des opérations et prend les décisions. Les différents acteurs ne sont impliqués qu'à titre de source d'information, et cela, à des moments particuliers de la démarche. Seul l'évaluateur a l'autorité pour juger de la valeur et du mérite du programme.

L'*approche axée sur le développement* d'un programme est moins connue et d'un emploi moins répandu (Patton, 1994 ; Ridde *et al.*, 2006). Son intérêt est que les résultats et le processus même de l'évaluation sont utilisés pour faire évoluer le programme au cours de sa mise en œuvre. L'expert en évaluation guide les parties prenantes pour faire en sorte que le projet s'améliore, s'adapte et réponde mieux aux besoins des participants. Il est membre à part entière de l'équipe du projet et participe aux prises de décisions, tant sur le plan de l'évaluation que sur celui du développement du projet.

Quant à l'*approche participative*, elle préconise la mise en œuvre du processus évaluatif selon une démarche de coopération entre des évaluateurs

et des personnes concernées par l'évaluation, mais qui ne sont pas des évaluateurs professionnels. Trois raisons peuvent justifier le recours à une telle approche: la première est d'ordre pragmatique (accroître les chances de réussite du processus et maximiser l'utilisation des résultats), la deuxième est politique (souci démocratique de participation citoyenne et d'émancipation), et la troisième est épistémologique (il n'existe pas de réalité unique, et la pluralité des points de vue est essentielle à la construction d'un argument) (Weaver et Cousins, 2005). On peut distinguer deux tendances au sein de l'approche participative:

L'évaluation pratique sert à résoudre des problèmes, à améliorer un programme et à prendre des décisions. Elle est réalisée dans le cadre d'un partenariat égalitaire entre l'évaluateur et les parties prenantes, ces dernières étant essentiellement les personnes devant prendre les décisions.

L'évaluation émancipatrice est un moyen, mais aussi, dans une certaine mesure, une fin en soi. Le processus évaluatif doit favoriser l'accroissement du pouvoir d'agir (*empowerment*) des parties prenantes, dans une perspective de justice sociale. Les décisions à l'égard du processus évaluatif sont exclusivement prises par les parties prenantes, et non par l'évaluateur (Cousins et Whitmore, 1998; Ridde, 2006).

Les normes de pratique

À la suite de la création de multiples associations professionnelles et regroupements nationaux se consacrant à l'évaluation de programmes, essentiellement dans les pays à revenus élevés, l'année 2003 a vu naître l'Organisation internationale de coopération en évaluation (OICÉ). Dans les pays à moyens et faibles revenus, de telles associations sont encore rares. Dans l'Afrique de l'Ouest francophone, par exemple, l'une des premières à voir le jour a été le Réseau nigérien de suivi et évaluation (ReNSE), créé dans la foulée de la première conférence de l'Association africaine d'évaluation (AfrEA), tenue en 1999. L'une des premières tâches de ces associations professionnelles d'évaluateurs a été de se doter de normes de pratique, de codes d'éthique ou, plus récemment, d'une liste des compétences essentielles propres à cette profession. L'Association américaine d'évaluation, par exemple, a produit une liste de 30 critères garantissant la qualité d'une évaluation, critères qui sont répertoriés selon quatre dimensions clés d'une évaluation de programme de qualité: l'uti-

lité, la faisabilité, la propriété et la précision. L'AfrEA a tenté d'adapter ces principes (Rouge, 2004). Les Suisses ont remplacé la troisième dimension par la déontologie. Les Français et leur jeune association professionnelle d'évaluation, qui date de 1999, se sont munis de sept principes directeurs : pluralité, distanciation, compétence, respect des personnes, transparence, opportunité et responsabilité. Enfin, depuis peu, tant aux États-Unis qu'au Québec, des tentatives ont cours pour établir une liste des compétences essentielles que devraient détenir les évaluateurs (Doré et Marceau, 2006 ; Stevahn *et al.*, 2005 ; voir aussi en encadré, p. 201).

L'objectif de cette introduction n'était pas de dresser un portrait exhaustif du champ de l'évaluation de programme. En effet, il existe de multiples manières d'approcher ce champ de pratique et de définir les termes qui y sont employés. Nous espérons cependant avoir initié le lecteur aux concepts essentiels à la compréhension de l'évaluation de programme. Les chapitres de la première partie de l'ouvrage approfondissent certaines notions abordées nécessairement trop rapidement ici. Dans la seconde partie, des études de cas détaillées illustrent des applications empiriques de ces concepts et montrent notamment la richesse et la diversité des pratiques évaluatives.