

UNIVERSITE DE LIEGE

FACULTE DE MEDECINE

**ANALYSE DES PRATIQUES
D'EVALUATION DE PROGRAMMES
D'EDUCATION POUR LA SANTE**

**LES DEMANDES DE FINANCEMENT
A LA COMMUNAUTE FRANÇAISE DE BELGIQUE
EN 1992 ET 1993**

TOME I

THESE DE DOCTORAT

EN SANTE PUBLIQUE

PRESENTEE PAR MICHEL DEMARTEAU

ANNEE ACADEMIQUE 1997-1998

Ce travail est le fruit d'une longue collaboration entre des personnes qui se sont rencontrées et qui ont collaboré au sein de l'Association pour la Promotion de l'Éducation pour la Santé (APES) sur le thème de l'évaluation de l'éducation pour la santé.

Les fondateurs de cette association, ses administrateurs et sa cellule interuniversitaire ont perçu dès 1980 le besoin de développer la qualité des programmes d'action en Communauté française, en mettant à la disposition des acteurs la réflexion des équipes de recherche en matière d'évaluation. Je remercie particulièrement le professeur Gilberte Reginster-Haneuse, pour le soutien qu'elle m'a apporté en tant que présidente de l'APES, mais aussi pour la confiance qu'elle m'a accordée sur les plans humain et scientifique au moment de concrétiser cette thèse. Je remercierai également le professeur Dieudonné Leclercq, pour avoir assuré le secrétariat de cette association et pour ses interventions scientifiques toujours appréciées. Je remercierai enfin le trésorier de l'APES, Monsieur Albert Schleiper, auprès de qui j'ai beaucoup appris sur les aspects institutionnels et financiers, qui sont le quotidien des programmes d'action.

Au terme de ce parcours, que l'on pourrait qualifier à tort de personnel, mes remerciements s'adressent tout spécialement :

- *aux acteurs de l'éducation pour la santé en Communauté française de Belgique, qui sont à l'origine de ce travail et qui, je l'espère, en seront aussi les bénéficiaires ;*
- *aux chercheurs qui, depuis les débuts de l'APES ou à l'un ou l'autre moment de son existence, ont formé une équipe de travail stimulante pour répondre réellement aux besoins en évaluation des équipes de terrain sans leur imposer leur vision de l'évaluation ;*
- *et bien sûr à l'équipe actuelle, scientifique et administrative, qui m'a soutenu directement, en assurant un regard critique sur le fond et sur la forme de la thèse, et indirectement, en prenant à sa charge les nombreux travaux que je ne pouvais plus assurer : le professeur Philippe Mairiaux, Anne-Marie Degey, Katia Gagliardi, Sophie Grignard, Xavier Lechien, Nadia Mawet, Marie-Christine Miermans, Anne Poumay, Chantal Vandoorne.*

Je remercie particulièrement :

- *les services du gouvernement de la Communauté française de Belgique, en la personne de Monsieur W. Brunson, Directeur général de la Santé, pour m'avoir accordé l'accès aux dossiers de demandes de subvention ;*
- *les responsables des programmes que j'ai analysés, qui ont accepté de me rencontrer longuement pour raconter leur projet et leur manière d'envisager l'évaluation ;*
- *les membres de mon comité de thèse qui ont accepté de lire et de critiquer des documents qui n'étaient souvent que des ébauches : les professeurs M. Anseau, Président, J.-P. Deschamps, O. Kuty, D. Leclercq, Ph. Mairiaux, G. Reginster-Haneuse.*

Je voudrais, enfin, remercier les membres de ma famille : mon épouse, mes enfants et leurs grands-parents. Ils ont le plus directement soutenu ce travail qui leur a souvent paru très long. Je pense particulièrement à Chantal, qui a autant que moi contribué à cette thèse, en prenant à sa charge notre famille, mes travaux au sein de l'équipe de l'APES, la lecture et la correction du texte et l'indispensable soutien affectif.

SOMMAIRE

TOME I : LE CADRE THEORIQUE

INTRODUCTION

CHAPITRE 1 : LA GRILLE APEP : ANALYSE DES PRATIQUES D'EVALUATION DE PROGRAMMES

CHAPITRE 2 : LES MODELES ET PARADIGMES D'EVALUATION

CHAPITRE 3 : LES RECHERCHES SUR L'UTILISATION ET LA PRATIQUE DE L'EVALUATION

CHAPITRE 4 : L'EVALUATION DE L'EFFICACITE ET DE LA QUALITE EN SANTE PUBLIQUE ET EN PROMOTION DE LA SANTE

BIBLIOGRAPHIE

TOME II : LA RECHERCHE

CHAPITRE 5 : LE PROTOCOLE DE L'ANALYSE DESCRIPTIVE DES PRATIQUES D'EVALUATION DES PROGRAMMES D'EDUCATION POUR LA SANTE

CHAPITRE 6 : LE CONTEXTE DES PROGRAMMES

CHAPITRE 7 : LA STRUCTURE ET LE CONTENU DES PROGRAMMES

CHAPITRE 8 : LES MODALITES ET LES TYPES D'EVALUATION DES PROGRAMMES

CONCLUSIONS GENERALES

TABLES DES MATIERES DES TOMES I ET II

ANNEXE : LISTE DES ARGUMENTS ET DES PROPOSITIONS D'EVALUATION PAR PROGRAMME

INTRODUCTION

Tout évaluateur intervenant dans le cadre d'un programme d'action est rapidement confronté à deux problèmes : la diversité des modèles d'intervention disponibles et les enjeux de l'évaluation pour les différents acteurs du programme. En d'autres mots, il doit tout d'abord clarifier ce qu'est l'évaluation, sa nature, ainsi que le rôle qu'elle joue vis-à-vis de l'action, sa ou ses fonctions. En effet, l'évaluation ne couvre apparemment pas un ensemble d'activités homogènes.

La terminologie utilisée dans le champ de l'évaluation est particulièrement variée et disparate, mêlant les termes techniques et les approches théoriques, eux-mêmes reflètes d'enjeux méthodologiques et sociaux. Que signifient et à quoi servent donc : l'évaluation répondante, l'évaluation illuminante, l'évaluation ex post ou ex ante, l'évaluation de besoins ou l'évaluation de processus, l'évaluation formative ou sommative, l'évaluation transactionnelle, l'évaluation d'efficacité et l'assurance de qualité, l'évaluation diagnostique et l'évaluation indépendante des objectifs, l'évaluation scientifique, la recherche évaluative, l'évaluation orientée vers le client/le consommateur/le décideur, l'évaluation démocratique, l'évaluation de l'*input* ou des *outcomes*, l'évaluation stratégique, l'évaluation de l'efficacité, l'évaluation coût - avantage, l'auto-évaluation, etc.

A supposer qu'elle soit possible, la clarification de cette terminologie n'est pas l'objet de ce travail. Un traité ou un dictionnaire de l'évaluation serait plus adéquat. Cette pléthore terminologique laisse cependant perplexe. Est-il possible de décrire les activités d'évaluation dans le champ de l'éducation pour la santé et de la promotion de la santé autrement qu'en utilisant ces termes ? Faut-il décrire les méthodes, les techniques utilisées et/ou rendre compte des enjeux sociaux ?

Le premier problème est bien de découvrir la nature de l'évaluation, ses enjeux. Y a-t-il une évaluation ou des évaluations ?

Le second problème est celui de la spécificité ou non du champ de l'éducation pour la santé. Pose-t-il des questions particulières ? Comment y répondre en tenant compte de notre expérience de chercheur et d'acteur de l'éducation pour la santé, de nos intérêts, de nos compétences, mais aussi de nos limites.

L'introduction présente ces deux problèmes : les enjeux de l'évaluation et l'origine de notre recherche. Elle expose également la réponse que nous tentons d'apporter en précisant l'objectif et la méthode de travail ainsi que la structure générale du texte.

ENJEUX DE L'EVALUATION

L'évaluation est une démarche centrale dans tout processus d'action. On la trouve présente dans toutes les disciplines, dans tous les secteurs de l'activité humaine : l'éducation, l'action sociale, l'économie, la pratique scientifique, la pratique médicale, la santé publique... et l'éducation pour la santé.

Mais faire ce constat ne suffit pas. Il faut tenter de comprendre les raisons de cette apparente universalité. Qu'est-ce qui lie l'évaluation à l'action ?

L'évaluation semble jouer, vis-à-vis de l'action, trois fonctions différentes et complémentaires : connaître, juger et décider. En nous appuyant sur ce constat, ainsi que sur les réflexions théoriques de Barbier (1990, 1991) et Stufflebeam (1980), nous proposons au chapitre 1 un modèle descriptif des activités d'évaluation de programmes dans le champ de l'éducation pour la santé.

Pour comprendre l'évaluation, il faudrait également pouvoir développer une réflexion approfondie et interdisciplinaire sur ces trois fonctions et leurs rapports réciproques. Tâche importante, essentielle, qui n'est cependant pas à la portée d'un travail individuel et ponctuel. Nous insistons simplement sur l'importance de prendre en compte cette triple appartenance pour approcher les enjeux de l'évaluation.

L'EVALUATION SERT A CONNAITRE L'ACTION

L'évaluation pose de façon particulière la question des relations entre la recherche et l'action. Il ne faut pas la confondre avec la recherche-action, qui serait plutôt une méthode, voire un paradigme de recherche, ou avec la recherche sur l'action, qui identifie l'action comme un objet de recherche, ainsi que l'ont développé dans leurs travaux les sociologues comme Crozier et Friedberg (1977) ou Touraine (1966).

L'évaluation n'étudie pas l'action pour le savoir, mais pour l'action. Le savoir sur l'action n'est qu'une étape et non une finalité. Ainsi, l'évaluation n'est pas une science. Cette affirmation, qui soulève des enjeux épistémologiques importants, mériterait davantage de développements. L'évaluation sert-elle, par exemple, à accumuler des connaissances ? Des épistémologues, des sociologues, des philosophes et des historiens des sciences, comme Weber (1959), Latour (1991) ou Stengers (1997) développent de nombreux travaux sur les relations entre les sciences et l'action. Ces travaux permettent d'enrichir la réflexion sur les enjeux de l'évaluation. Comme les sciences, mais de manière plus aiguë encore, l'évaluation pose la question de la relation entre savoir et pouvoir.

Quant au lien de l'évaluation avec la recherche, il est aussi particulièrement évident sur le plan de la méthodologie. A cet égard, rien ne ressemble plus à une recherche qu'une évaluation dite "scientifique" ou recherche évaluative. On y retrouve, par exemple, les mêmes débats sur l'intérêt des méthodes qualitatives et quantitatives (Alkin, 1990). Cette dimension méthodologique, technique, doit être prise en compte pour décrire les pratiques d'évaluation en éducation pour la santé même si elle n'en constitue qu'une facette.

En réalisant une recherche sur l'évaluation, nous devons enfin éviter l'écueil d'un amalgame entre méta-évaluation et recherche sur l'évaluation. Notre étude est un projet de connaissance de l'évaluation et non de jugement des pratiques d'évaluation, et encore moins de modèle d'action pour l'évaluation.

L'EVALUATION SERT A JUGER L'ACTION

L'évaluation pose la question de la valeur des actions entreprises. C'est pour cela qu'elle est un enjeu stratégique entre les acteurs d'un programme d'action. C'est pour cela aussi qu'elle devient omniprésente dans le discours politique sous la pression de choix économiques et sociaux de plus en plus difficiles à poser. Le domaine de la santé sert d'exemple édifiant. Jamais on n'y a autant parlé de rendement, d'efficacité ou de contrôle de qualité. C'est la notion de "*accountability*", omniprésente dans la littérature anglo-saxonne.

Mais cette croissance de la demande évaluative ne peut se comprendre que par son rapport au savoir, à la connaissance, à la dictature de la science et des experts, comme n'hésite pas à l'affirmer Stengers (1997). Cette tendance procède d'une confusion, ou à tout le moins d'un manque de clarification, entre les notions de valeur et de vérité. La philosophie a depuis longtemps traité cette relation de multiples manières (Comte-Sponville). Un apport de la philosophie sur ce sujet serait aussi nécessaire.

Quoi qu'il en soit, le jugement de valeur est une composante essentielle de l'évaluation qui doit être intégrée dans tout modèle sur ce sujet.

L'EVALUATION SERT A GUIDER L'ACTION

La dernière facette de l'évaluation, mais aussi la plus importante, est la question de la décision qui conduit l'action. C'est la finalité ou l'usage de l'évaluation.

Cependant, si l'évaluation a pour but essentiel d'assurer la régulation de l'action (Leclercq, 1995, Stufflebeam, 1980), c'est parce qu'elle aide à penser et à juger l'action. Elle permet de construire une représentation de l'action (Barbier, 1991). La fonction de régulation est donc intimement liée aux deux précédentes.

LA TRIPLE FONCTION DE L'EVALUATION

En conclusion, il importe que toute tentative pour décrire des activités d'évaluation considère ces trois composantes. L'évaluation est une pratique sociale qui peut être objet de recherche, mais c'est un objet particulier à la croisée de la science, de la philosophie et du politique. Idéalement, cette recherche sur l'évaluation devrait intégrer l'apport de réflexions scientifiques sur les connaissances, sur les valeurs et sur l'action.

Est-il possible d'identifier et de décrire toutes les formes d'évaluation de programmes d'éducation pour la santé dans un modèle d'analyse unique reprenant ces trois éléments ?

La littérature nous apporte des arguments pour différencier les pratiques sur la base de ces trois dimensions tout en conservant l'idée d'une unité de structure à toute évaluation. La combinaison de ces trois axes permettrait de couvrir les différentes pratiques d'évaluation.

Alkin et Ellett (1990) ont comparé différents modèles proposés par des théoriciens de l'évaluation en recherchant les principes sous-jacents. Par "modèle", ils font référence à un ensemble large et cohérent d'idées développées par un auteur à un moment donné. Ils reconnaissent toutefois la difficulté d'identifier comme "modèle" des textes qui proposent soit des concepts généraux, soit des démarches à suivre, qui évoluent dans le temps et qui omettent de préciser ou d'argumenter leurs hypothèses de travail.

Alkin et Ellett proposent trois dimensions pour caractériser et différencier les modèles d'évaluation : la méthodologie, la définition (le jugement de valeur) et les fonctions (l'usage de l'évaluation) attribuées à l'évaluation.

La méthodologie permet de décrire et d'expliquer les phénomènes étudiés. Les débats entre méthodes expérimentales, quasi-expérimentales et non expérimentales, entre données quantitatives et qualitatives font partie de ce champ de réflexion.

La définition donnée à l'évaluation précise la place du jugement de valeur dans un modèle, la manière d'attribuer une valeur à un objet. Le débat porte sur le rôle de l'évaluateur et sa position vis-à-vis du jugement de valeur : éducateur, conseiller, arbitre pour des décideurs ou vis-à-vis de la société.

Les fonctions permettent de préciser l'usage de l'évaluation pour des décideurs, des utilisateurs, en interaction ou non avec ceux-ci.

Ces auteurs font donc apparaître les trois composantes que nous avons présentées ci-dessus. L'originalité de leur démarche est d'avoir montré que chaque modèle présentait ces trois dimensions tout en accentuant l'une ou l'autre d'entre elles.

Alkin et Ellett plaident pour une approche plus descriptive, plus empirique de l'évaluation et pour l'étude du processus de développement des modèles d'évaluation.

Leur démarche a inspiré notre travail : construire un modèle d'analyse qui rende compte de toutes les pratiques d'évaluation de programmes dans un champ et un contexte particuliers.

ORIGINE DE LA RECHERCHE

Une recherche naît dans un contexte particulier qui l'oriente et qu'il est nécessaire de présenter pour en faciliter la compréhension et la critique. Les situations choisies pour l'analyse, les concepts, les options méthodologiques et les outils trouvent leur origine, leurs fondements dans une approche "historique" du parcours personnel du chercheur et du développement de l'éducation pour la santé en Communauté française de Belgique.

Plus précisément, il semble particulièrement opportun de clarifier notre position de chercheur et d'acteur dans le secteur de l'éducation pour la santé pour situer les raisons de notre intérêt pour une "analyse des pratiques d'évaluation en éducation pour la santé".

De ce contexte émergeront les questions qui sont à la base de notre travail.

LES PREMIERS PAS D'UN PARCOURS SCIENTIFIQUE

Tout d'abord comme étudiant en éducation physique (Demarteau et Pieron, 1978), puis en sciences de l'éducation, nous travaillons sur la problématique de l'analyse des processus d'enseignement. Nous utilisons à cette occasion des grilles d'observation et d'analyse.

Comme chercheur en sciences de l'éducation, dès 1980, nous approfondissons les concepts d'évaluation prédictive, formative, et sommative : construire des outils d'évaluation, former des enseignants, étudier leurs attitudes sur ces concepts (Demarteau et Simenya, 1982, Simenya et Demarteau, 1984).

L'EXPERIENCE EN EDUCATION POUR LA SANTE EN COMMUNAUTE FRANÇAISE

A partir de 1982, nous assurons la coordination d'une cellule interuniversitaire de conseil en méthodologie et en évaluation chargée de développer les contacts entre des équipes de terrain et des services universitaires, puis, progressivement, nous effectuons nous-même l'activité d'évaluateur et de conseiller en évaluation (Demarteau, 1984, Demarteau et Reginster-Haneuse, 1988, Demarteau et Reginster-Haneuse, 1988, Vandoorne, Demarteau et Kefer, 1992, Demarteau et Vandoorne, 1994).

Cette recherche s'inscrit enfin dans notre travail de formateur en éducation pour la santé au sein de l'école de santé publique de l'Université de Liège. La planification et l'évaluation des actions d'éducation pour la santé est un axe de cette formation (Demarteau, 1992b). Comme formateur, nous prescrivons des démarches de planification et d'évaluation, mais en nous basant plutôt sur notre expérience et certains modèles théoriques que sur les résultats d'une recherche.

Supprimé :

Cette base théorique issue de notre formation en éducation physique, en sciences de l'éducation et en santé publique, mais aussi le cumul des expériences d'une quinzaine d'années de pratique d'expertise et de formation en éducation pour la santé expliquent notre intérêt pour le thème de cette thèse et les options qui y sont reflétées.

LA PRATIQUE DU CONSEIL EN EVALUATION

Au centre de nos préoccupations au moment de débiter cette étude, se situe la fonction de conseiller en évaluation (Demarteau, 1992c, Demarteau, 1994) exercée dans le cadre d'un service agréé pour cette mission par la Communauté française de Belgique. Il y a là, nous semble-t-il, un enjeu pour la qualité des actions d'éducation pour la santé.

Le conseiller en évaluation n'évalue pas un programme, il le soutient. Il aide les acteurs de celui-ci à concevoir, mettre en oeuvre et exploiter leurs propres évaluations. En paraphrasant la définition de l'évaluation proposée par Stufflebeam (1980), on pourrait dire que le conseil en évaluation est un processus qui vise à aider les acteurs d'un programme à délimiter, obtenir et exploiter les informations utiles pour prendre les décisions nécessaires à la conduite du programme.

L'évaluation interne au programme est une réalité : nos interactions avec les professionnels du terrain l'ont montré. Ainsi, les interventions, les conseils pour développer des programmes de qualité doivent s'appuyer sur les pratiques d'évaluation déjà existantes. Mais ces pratiques sont différentes de celles des évaluateurs "professionnels". La pertinence, la qualité, l'efficacité des conseils en évaluation et leurs effets sur la qualité et l'efficacité des programmes passent par une meilleure description des pratiques d'évaluation menées par les promoteurs de programmes eux-mêmes. C'est l'origine principale de cette thèse.

En quoi et pourquoi les pratiques d'évaluation des acteurs sur le terrain sont-elles différentes des modèles théoriques ? Le conseil en évaluation est-il pertinent, efficace ? Identifier ces pratiques et proposer des modèles d'interventions en évaluation mieux adaptés aux besoins et aux demandes du terrain apporterait une contribution au développement et à l'efficacité de l'éducation pour la santé.

Cette histoire, ces enjeux nous ont conduit à investir d'abord dans un approfondissement théorique, puis dans une recherche sur les pratiques d'évaluation de programmes en éducation pour la santé.

UN APPORT ORIGINAL

Si l'évaluation est un domaine théorique fort vaste, par contre la recherche sur l'évaluation et plus particulièrement l'étude des pratiques d'évaluation, surtout non professionnelles, couvre un champ beaucoup plus limité. Dans le domaine de l'éducation pour la santé en particulier, cette étude investit une problématique peu investiguée bien que présente dans les préoccupations de nombreux praticiens et décideurs.

Malgré une revue de littérature très large, abordant d'autres secteurs de l'action sociale comme l'éducation, le travail social, la santé publique, nous n'avons pu identifier une "théorie" générale de l'évaluation de programmes capable de rendre compte de l'ensemble des "modèles" construits par les évaluateurs et d'intégrer les pratiques d'évaluation interne que nous rencontrons sur le terrain.

Selon Alkin (1990), il n'y aurait d'ailleurs pas de théorie explicative propre à l'évaluation qui comprendrait un ensemble de propositions générales capables de décrire, expliquer et prédire toutes les activités d'évaluation.

Un nombre considérable de "modèles" d'évaluation sont des théorisations des pratiques de professionnels de l'évaluation, pour la plupart américains. Il existe des taxonomies de ces modèles, qui, le plus souvent, les juxtaposent plutôt que de les intégrer sur la base d'un principe commun sous-jacent. D'autres recherches décrivent les pratiques des professionnels de l'évaluation ou étudient les conditions d'utilisation des évaluations. Mais rares sont les propositions qui prennent en compte les trois fonctions que nous identifions dans l'évaluation. Seuls l'article d'Alkin et Ellett (1990) et le cadre théorique de Barbier (1990) proposent ces trois composantes d'une évaluation : connaître, juger, décider.

OBJECTIF ET METHODE DE LA RECHERCHE

Après avoir abordé brièvement les enjeux de l'évaluation et l'origine de notre travail, nous terminerons cette introduction en présentant la démarche que nous avons élaborée et la structure de notre texte.

QUESTIONS DE DEPART

Au départ de notre travail, nous souhaitons répondre à trois groupes de questions qui correspondent aux trois phases d'une recherche (décrire, expliquer, expérimenter) :

1. Quelles sont les pratiques d'évaluation de programmes en éducation pour la santé ?

Cette première phase est descriptive. Qui fait de l'évaluation et comment ? Qui parle d'évaluation, dans quels termes, en référence à quels modèles ? Qui demande, exige l'évaluation et pourquoi ? L'investigation a été limitée aux pratiques actuelles d'évaluation des programmes d'éducation pour la santé dans la Communauté française de Belgique.

2. Qu'est ce qui détermine ces pratiques ?

Cette phase est explicative. Qu'est-ce qui influence les pratiques ? Le contexte politique, économique, législatif ? Les caractéristiques de la situation d'application du programme d'actions ? Le modèle d'intervention en éducation pour la santé (médical, social, éducatif) ? Les caractéristiques individuelles des acteurs et des concepteurs de programmes ? Leurs représentations de l'évaluation (sa difficulté, son utilité) ? Leur statut professionnel dans leur organisation et dans la société ? Leur formation professionnelle ? Les caractéristiques des évaluateurs externes au programme ?

3. Quelles sont les pratiques d'évaluation qui sont efficaces pour le développement des actions d'éducation pour la santé ?

Cette phase, qui devrait être expérimentale, représente un prolongement potentiel de l'étude actuelle. Elle vérifierait si certaines pratiques sont plus utiles, plus efficaces, pour qui, à quelles conditions, sous quelles formes.

Les pratiques d'évaluation déterminent-elles des effets différents sur la gestion d'un programme, sur les résultats immédiats de celui-ci, sur l'état de santé de la population touchée ? Le conseil en évaluation peut-il modifier des pratiques et dans quel sens, avec quels bénéfices ?

OBJECTIF DE LA RECHERCHE

Le but de l'étude est donc de construire un modèle qui permette de décrire et d'expliquer toutes les pratiques d'évaluation en éducation pour la santé susceptibles d'être rencontrées dans le contexte des programmes financés par les pouvoirs publics.

Objet de l'étude

Pour atteindre ce but, nous avons tout d'abord recherché dans la littérature des éléments nous permettant d'identifier et de décrire notre objet d'étude : la pratique des évaluations de programmes d'éducation pour la santé.

Deux termes méritent d'emblée d'être précisés : pratique et programme. Ils situent le contexte de notre réflexion.

Par pratique (Petit Robert, 1976), nous entendons *"la manière concrète d'exercer une activité d'évaluation par opposition à une manière abstraite et spéculative (théorie)"*. La pratique d'évaluation s'exerce dans un cadre professionnel qui est celui de la conception et de la gestion d'un programme d'éducation pour la santé. Mais cette pratique n'est pas nécessairement le fait d'un professionnel de l'évaluation. Elle peut être observée et décrite. Si elle intègre la représentation de l'évaluation, elle la dépasse en s'exprimant dans des actes. Elle est déterminée aussi bien par des contraintes techniques propres à l'évaluation que par des contraintes sociales.

Dans notre étude, ces pratiques seront identifiées à partir des documents qui font état de la conception et de la gestion des programmes : le projet écrit et le rapport écrit. Leur description sera enrichie du contenu des entretiens avec les gestionnaires de programmes.

Par programme (Petit Robert, 1976), nous entendons *"une suite d'actions que l'on se propose d'accomplir pour arriver à un résultat"* et que l'on accomplit. Le programme n'est pas une *"suite formalisée d'opérations"* (autre sens du mot). Il est constitué d'un projet et d'une démarche de gestion de ce projet au sens proposé par Barbier (1991). Il ouvre un espace de gestion plus autonome à un promoteur, *"privilégiant les tâches de conduite par rapport aux tâches de réalisation"* (Barbier, 1991, p. 18). Barbier distingue également l'évaluation des programmes de l'évaluation des personnes (Barbier, 1990).

Démarche méthodologique

A partir de notre objectif initial, notre premier travail a été de construire une grille d'analyse générale des pratiques d'évaluation des programmes d'éducation pour la santé. L'ébauche en quelque sorte de la théorie générale que nous recherchions et qui permette de faire un lien entre les pratiques professionnelles décrites par la littérature nord-américaine et les pratiques de terrain des acteurs des programmes.

Cette grille, ses fondements théoriques et la formulation de la thèse constituent le premier chapitre de notre texte.

La validation théorique de cette grille est présentée dans les trois chapitres suivants, qui la confrontent aux pratiques proposées dans la littérature (chapitre 2), aux recherches menées sur l'évaluation (chapitre 3) et à quelques questions récurrentes rencontrées dans le champ de la santé, de la santé publique, de la promotion de la santé et de l'éducation pour la santé (chapitre 4).

La deuxième étape constitue la recherche sur les activités d'évaluation proposées par des promoteurs de programme. Nous souhaitons vérifier la présence ou non, au sein des programmes, des différentes formes d'évaluation prévues par le modèle et identifier des facteurs en relation avec leur apparition.

Le chapitre 5 présente la méthodologie de recherche et les hypothèses de départ.

Les trois chapitres suivants détaillent les résultats et les discutent. Nous identifierons le contexte des programmes en décrivant les caractéristiques des promoteurs, des interventions proposées et du protocole d'évaluation (chapitre 6). Nous analyserons ensuite le contenu des programmes en relation avec le contexte (chapitre 7). Nous proposerons enfin, sur base de la grille d'analyse, une description des modalités d'évaluation effectivement rencontrées ; nous montrerons également les liens entre les caractéristiques de contenu et de contexte et certaines de ces pratiques d'évaluation (chapitre 8).

Ce travail est donc une recherche exploratoire réalisée à l'aide d'une grille d'analyse très globale, mais avec un champ d'application limité dans le temps et dans l'espace. Ce modèle permet-il de décrire les pratiques d'évaluation des programmes développés en Communauté française de Belgique durant les deux années qui ont précédé le début de l'étude. Ces pratiques sont-elles diversifiées ? Reflètent-elles les différentes fonctions que peut prendre une évaluation ? Varient-elles avec le contenu ou le contexte des programmes ?

CHAPITRE 1

LA GRILLE APEP :

ANALYSE DES PRATIQUES

D'EVALUATION DE PROGRAMMES

I.	UNE DEFINITION DE L'EVALUATION.....	16
II.	LES AXES DE DIFFERENCIATION DES PRATIQUES D'EVALUATION.....	18
II.A.	LA DIMENSION SOCIALE OU L'AXE IMPLICITE - EXPLICITE.....	18
II.B.	LA DIMENSION STRATEGIQUE OU L'AXE INTERNE – EXTERNE.....	20
II.C.	LA DIMENSION TECHNIQUE OU L'AXE INFORMEL – FORMEL.....	29
III.	LA GRILLE D'ANALYSE APEP.....	34
III.A.	LES HUIT TYPES D'EVALUATION.....	34
III.B.	LES 124 MODALITES D'EVALUATION DE LA GRILLE D'ANALYSE.....	37
IV.	LA DESCRIPTION DES PROGRAMMES.....	39
IV.A.	LA STRUCTURE D'UN PROGRAMME.....	40
IV.B.	LE CONTENU D'UN PROGRAMME.....	44
V.	LA THESE	47

L'élaboration de la grille théorique cherche à rencontrer deux exigences : tout d'abord, intégrer les modèles théoriques existants ; ensuite, permettre de décrire les pratiques d'évaluation en usage sur le terrain, aussi bien par des évaluateurs spécialisés que par des praticiens de l'éducation pour la santé.

Dans ce premier chapitre, nous présentons la grille de manière globale en essayant de refléter les discussions, débats, expériences, résultats de recherches qui fondent les choix opérés et l'opérationnalisation des catégories de la grille présentée plus tard dans le chapitre 5.

Pour construire celle-ci, nous avons pris comme point de départ une définition la plus large possible de l'évaluation en identifiant trois composantes à une activité d'évaluation : connaître, juger, décider.

Nous avons ensuite dégagé trois axes de différenciation des pratiques applicables à chacune de ces composantes :

- L'explicitation sociale ou non des composantes de l'évaluation. Ce principe sous-jacent permet de distinguer huit grands types d'évaluation. C'est la dimension sociale de l'évaluation.
- La position interne ou externe des acteurs intervenant dans chacune des trois composantes explicitées de l'évaluation. C'est la dimension stratégique de l'évaluation.
- Le caractère informel ou formel de la mise en œuvre des trois composantes explicitées de l'évaluation. C'est la dimension technique de l'évaluation.

La combinaison des trois composantes et des trois dimensions fournit une grille théorique qui prévoit l'existence potentielle de 124 modalités d'évaluation différentes.

La thèse que nous présentons soutient que cette grille permet aussi bien de décrire les pratiques d'évaluation de terrain en éducation pour la santé que les pratiques théorisées des professionnels de l'évaluation.

I. UNE DEFINITION DE L'EVALUATION

Nevo (1986), dans une revue analytique de la littérature sur l'évaluation en éducation, identifie trois grandes tendances dans les définitions de l'évaluation.

La première, la plus courante, définit l'évaluation comme une mesure de la valeur d'un objet, d'un programme, en référence à des auteurs comme Eisner, House, Scriven. On peut associer Barbier (1990) à cette tendance. Le terme évaluation est étymologiquement lié à la notion de jugement de valeur.

La seconde, moins fréquente, n'accepte pas le jugement de valeur et met l'accent sur la récolte d'informations. L'évaluation est une investigation systématique du déroulement et des conséquences d'un programme. Ce processus fait l'objet de l'attention méthodologique de professionnels de l'évaluation comme Rossi et Freeman (1979), Cronbach (1980), Suchman (présenté par Stufflebeam et Shinkfield, 1985), Weiss (dans un article repris dans l'ouvrage de Alkin, 1990).

Une variante de cette deuxième tendance consiste à intégrer dans la notion d'évaluation à la fois le jugement, produit de l'action d'évaluer, et le processus qui y conduit. L'évaluation est un processus de production d'un jugement de valeur. Nevo (1986) cite à l'appui de cette option des auteurs tels que Stake, Guba et Lincoln.

La troisième tendance définit l'évaluation comme le moyen de fournir de l'information pour prendre des décisions. Cette position est souvent défendue par des théoriciens de l'évaluation tels que Alkin (1979), Stufflebeam (1980). Ce dernier, en introduction d'un ouvrage publié en 1971, précisait que *"The purpose of the evaluation is not to prove but to improve"*. On peut associer à ce courant théorique la position de Leclercq (1995) qui donne prioritairement à l'évaluation un rôle de régulation.

Pour comprendre les pratiques d'évaluation, il faut donc envisager aussi l'usage du jugement de valeur comme le suggère Barbier (1990). Dans le cadre de l'évaluation de programmes, cet usage est la conduite de l'action, la prise de décision.

Nous définirons donc l'évaluation de programmes comme un processus de production d'un jugement de valeur pour conduire une action par des décisions.

Une évaluation comporte donc trois phases ou trois composantes qui sont le processus, le produit et son usage.

Le processus est l'acte d'évaluation. Il vise à délimiter, obtenir et fournir des informations sur la valeur d'un objet ou d'une action (Stufflebeam, 1980). Dans une première étape, il identifie des objets d'évaluation et précise des critères de jugement. Dans la seconde, il décrit les objets d'évaluation en utilisant des techniques de recueil d'information et il les interprète en fonction des critères. Dans la troisième, il met en place des moyens et des procédures pour communiquer le produit de l'évaluation. Nous tenterons d'identifier ce processus dans nos observations.

Le produit de l'évaluation est un jugement de valeur. C'est parce qu'elle produit des jugements de valeur que l'évaluation est un enjeu pour les acteurs d'un programme. Nous tenterons d'identifier ce produit et les enjeux qu'il soulève.

L'utilisation de l'évaluation pose la question des objectifs et des effets de l'évaluation. Nous nous intéresserons aux utilisations de l'évaluation et à ses effets sur la conduite des programmes. Plus particulièrement, nous tenterons de situer le rôle de l'évaluation en regard des quatre grandes catégories de décision identifiées par Stufflebeam (1980) dans son modèle CIPP (Contexte, Intran, Processus, Produit) : la formulation des objectifs, le choix des stratégies d'intervention et des ressources, la régulation du programme pendant son déroulement et les décisions sur la poursuite de celui-ci.

II. LES AXES DE DIFFERENCIATION DES PRATIQUES D'EVALUATION

Entre une définition de l'évaluation et une description des pratiques, il reste toutefois un long chemin à parcourir. Si la définition postule une unité de nature à l'évaluation, comment expliquer la diversité et parfois l'opposition manifestée entre certaines pratiques théorisées. S'agit-il de différences apparentes, et il n'y aurait qu'une et une seule évaluation, ou de différences réelles qui justifieraient que l'on parle des évaluations au pluriel comme le suggère Monnier (1987) ?

La littérature apporte des arguments pour identifier trois dimensions importantes qui différencient les pratiques tout en conservant l'idée d'une unité de structure à toute évaluation. La combinaison de ces trois axes de différenciation permettrait de couvrir l'ensemble des pratiques d'évaluation.

II.A. LA DIMENSION SOCIALE OU L'AXE IMPLICITE - EXPLICITE

L'évaluation est une pratique sociale. Cette affirmation, pour évidente qu'elle paraisse, a mis longtemps à s'imposer à la majorité des évaluateurs (Cronbach, 1980, Walberg, 1990, Sensi, 1994). Elle ne fait d'ailleurs toujours pas l'unanimité (Weiss, 1990).

Cette pratique sociale de l'évaluation s'exprime, en première analyse, par la position sociale et les intérêts des différents acteurs de l'évaluation. Sensi (1994) met bien en évidence cette dimension politico-stratégique de l'évaluation, illustrée dans la présentation du deuxième axe de différenciation.

Mais de façon plus générale, et plus subtile aussi, la dimension sociale s'exprime par la mise en évidence, la socialisation, de l'une ou l'autre des composantes de l'évaluation. L'ouvrage de Barbier sur l'évaluation en formation, et plus particulièrement le chapitre sur les "formes sociales d'émergence" (1990, pp. 28 à 36) ouvre la voie à l'introduction du concept d'explicitation sociale.

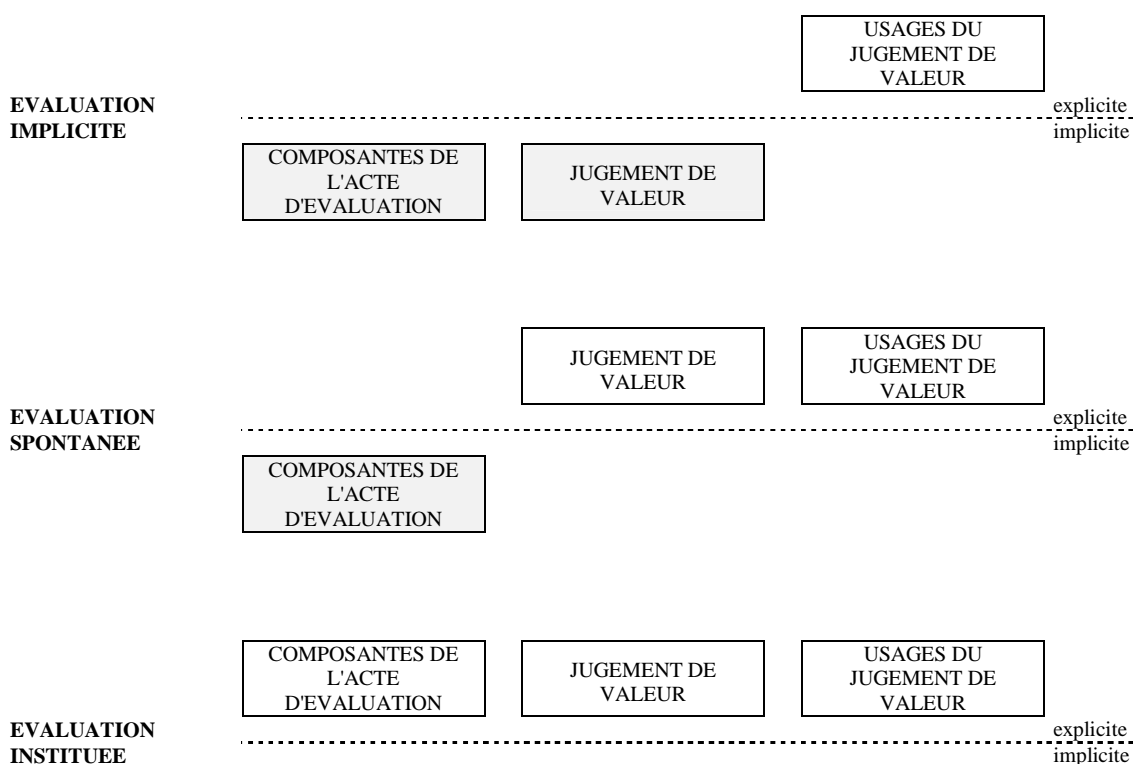
Dans ce chapitre, Barbier s'efforce de réaliser *"un premier inventaire de ces pratiques qui précisément se désignent comme pratiques d'évaluation"* (p. 27). Il constate que *"d'une manière générale tout ce qui se dit ou s'écrit sur le déroulement concret des activités de formation tend actuellement à être désigné sous le terme d'évaluation : évaluation bien sûr le compte rendu que le responsable d'un stage doit fournir à son commanditaire éventuel, évaluation aussi le contrôle qu'un gestionnaire de la formation veut exercer sur les coûts, évaluation encore l'étude qu'un chercheur peut conduire sur le fonctionnement d'un dispositif pédagogique très précis ou d'un système de formation de grande dimension. Il y a là bien évidemment un usage abusif de la notion..."* (p. 28).

A partir de ce constat, Barbier va s'efforcer de distinguer ce qui est ou n'est pas de l'évaluation, en s'appuyant comme d'autres (Stufflebeam et Webster, 1980) sur la production ou non d'un jugement de valeur. Il tente tout d'abord de clarifier la confusion entre les notions de contrôle, d'analyse et d'évaluation.

La définition par la présence d'un jugement de valeur lui apparaissant insuffisante pour désigner les phénomènes d'évaluation, Barbier entreprend de distinguer trois formes de manifestation du jugement de valeur : l'évaluation "implicite", "spontanée" et "instituée".

Comme le montre le tableau suivant, la différenciation de ces trois formes se fait sur la nature implicite ou explicite de trois éléments constitutifs de l'évaluation : les composantes de l'acte d'évaluation, le jugement de valeur et l'usage du jugement de valeur.

Formes de manifestation de l'évaluation en formation (Barbier, 1990, p. 35)



Pour Barbier, l'évaluation "implicite" est la plus fréquente. "*Le jugement de valeur ne s'explique qu'à travers ses effets*" (Barbier, 1990, p. 32). Cette première forme de manifestation de l'évaluation en fait, pour certains, un acte social universel. Toute perception d'autrui ou toute progression d'action suppose une évaluation de la valeur d'un individu ou d'une action. Ce caractère universel ne facilite pas l'identification de l'évaluation. Il peut cependant servir à en repérer les autres formes de manifestation.

Il y a évaluation "spontanée" lorsqu'un jugement de valeur est exprimé spontanément sur une activité ou une personne : *"le jugement de valeur ne s'explique qu'à travers son énoncé, sa formulation"* (Barbier, 1990, p. 33). C'est une forme extrêmement fréquente également, qui ne comprend cependant pas d'explicitation des critères. Ils ne sont pas informels, mais ils ne sont ni socialement organisés ni émis par des personnes socialement déléguées pour le faire. Ce type d'évaluation ne doit rien au hasard. Elle est souvent vécue comme un droit d'expression. Elle détermine des politiques par la succession des jugements, s'inscrivant dans un rapport complexe avec les évaluations plus structurées, qui servent parfois à la remplacer ou à la canaliser.

L'évaluation "instituée" est la dernière forme de manifestation de l'évaluation : *"le jugement de valeur s'explique entièrement dans sa production comme résultat d'un processus social spécifique dont les principales étapes sont susceptibles d'observation"* (Barbier, 1990, p. 34). *"C'est un acte délibéré et socialement organisé aboutissant à la production d'un jugement de valeur"* (Barbier, 1990, p. 34). Pour Barbier, l'évaluation "instituée" présente les caractéristiques suivantes : des instruments et une méthodologie, des personnes socialement habilitées, des résultats explicités pour être utilisés.

La grille théorique que nous proposons s'inspire des concepts développés par Barbier, mais en les généralisant.

Nous conservons la distinction entre le processus (les composantes de l'acte d'évaluation), le jugement (le jugement de valeur) et la décision (l'usage du jugement de valeur).

Nous gardons aussi l'axe de différenciation implicite - explicite, caractéristique de la socialisation des composantes de l'évaluation.

Mais nous divergeons avec Barbier sur un point important. Nous ne faisons pas du jugement de valeur la caractéristique essentielle, le pivot de l'évaluation. L'explicitation ne se définit pas exclusivement en référence avec cette composante, mais également avec les autres.

Pour nous, toutes les pratiques d'évaluation ont en commun un processus, un jugement de valeur et un usage de ce jugement. Ce qui les différencie c'est la fonction privilégiée que l'évaluateur attribue à une ou plusieurs de ces caractéristiques. C'est ainsi que les fonctions de l'évaluation sont liées, selon les auteurs, à la fonction privilégiée qu'ils lui attribuent : comprendre (processus), juger (produit) ou réguler (usage).

Il faut pouvoir intégrer dans un modèle descriptif général des pratiques d'évaluation les nombreuses formes d'évaluation qui ignorent ou rejettent le jugement de valeur. Il nous semble que ces cas relèvent d'une absence d'explicitation sociale, volontaire, involontaire ou différée, et non d'un usage abusif du terme évaluation comme l'affirment Barbier (1990) ou d'autres avant lui (Stufflebeam et Webster, 1980).

C'est pourquoi nous abandonnons aussi les dénominations "implicite", "spontané" et "institué". De très nombreuses évaluations sont socialement instituées, mais ne sont pas explicitées dans tous leurs éléments. C'est le cas de l'évaluation "scientifique", de l'expertise, de l'inspection. Ce sont toutes des pratiques d'évaluation instituées, réalisées par des personnes socialement habilitées.

En synthèse, nous pensons que les trois composantes de l'évaluation sont toujours présentes, mais que c'est l'énonciation d'une ou de plusieurs de ces composantes qui est significative de la fonction sociale qui est attribuée à l'évaluation et qui détermine sa pratique. Cette explicitation est pour nous le signe de la valeur sociale qui lui est attribuée par les acteurs. C'est l'élément déterminant qui permet de différencier les multiples pratiques d'évaluation.

L'évaluation est avant tout une pratique sociale. C'est cette pratique sociale qui permet d'interpréter les multiples modèles et pratiques rencontrés dans la littérature et sur le terrain. Cette pratique sociale peut être identifiée à partir de la reconnaissance du caractère implicite ou explicite des composantes de l'évaluation. Les caractéristiques des trois composantes de l'évaluation "explicitée" seront donc observables : des instruments et une méthodologie, des personnes socialement habilitées, des jugements et des décisions énoncées socialement.

L'erreur serait d'expliquer d'une part la nature du processus d'évaluation par une approche instrumentale sur les méthodes et les techniques ; d'autre part la nature du jugement et de son usage par une approche sociale. Ni procédure instrumentale, ni processus social, mais les deux à la fois, l'évaluation de programmes doit être étudiée comme une pratique sociale spécifique.

C'est le propos de Barbier (1990) que nous avons tenté d'élargir et d'appliquer au champ de l'éducation pour la santé. L'axe implicite - explicite appliqué aux trois composantes de l'évaluation permet de prévoir l'existence de huit types d'évaluation qui font ici l'objet d'une première présentation, mais qui seront détaillés plus loin dans le cadre de la présentation de la grille d'analyse et qui serviront ensuite de fil conducteur pour l'examen de la littérature et la phase de recherche.

Les huit types d'évaluation

T0 : L'évaluation implicite

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T1 : L'évaluation "décision" (décision explicite)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T2 : L'évaluation d'expertise (jugement explicite)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T3 : L'évaluation formelle (processus explicite)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T4 : L'évaluation de contrôle (jugement et décision explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T5 : L'évaluation orientée vers les conclusions (processus et jugement explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T6 : L'évaluation "processus - décision" (processus et décision explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

T7 : L'évaluation explicite (processus, jugement et décision explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

II.B. LA DIMENSION STRATEGIQUE OU L'AXE INTERNE - EXTERNE

Nous avons évoqué plus haut une dimension politico-stratégique de l'évaluation proposée par Sensi (1994). Dans la dimension politico-stratégique, Sensi identifie une série de questions sur les buts et les acteurs de l'évaluation : qui la demande et pourquoi ; qui participe, qui contrôle et comment. Elle justifie cette dimension par le principe démocratique proposé par MacDonald (cité par Sensi, 1994) et par l'analyse stratégique du pouvoir dans les organisations définie par Crozier et Friedberg (1977). Nous retiendrons cette dimension stratégique liée aux acteurs comme un élément distinct de la dimension sociale, bien qu'elle en fasse partie. Elle ne s'applique toutefois qu'aux composantes de l'évaluation déjà explicitées.

Dans leur ouvrage "L'acteur et le système", Crozier et Friedberg (1977) n'utilisent pas le terme "évaluation". Leur contribution est néanmoins essentielle pour comprendre la place que peut jouer l'évaluation dans une organisation.

L'évaluation comme source d'information, de jugement de valeur et de décision semble en effet jouer un rôle dans les jeux de pouvoir entre acteurs d'une action organisée (Crozier et Friedberg, 1977, pp. 91 à 127). C'est particulièrement le cas dans la formulation des objectifs d'une action, point clé d'une approche évaluative, mais, souvent aussi, point d'achoppement de celle-ci faute d'une analyse des enjeux sous-jacents. C'est aussi le cas dans la mesure des résultats de l'action au sein de l'organisation et dans son environnement. *"La mesure du résultat (...) offre le moyen d'instrumentaliser en les médiatisant une partie des processus politiques caractéristiques de l'action collective et, ce faisant, contribue à les rendre plus clairs et rationnels."* (Crozier et Friedberg, 1977, p. 176).

L'évaluation pourrait aussi intervenir dans le *"processus d'apprentissage collectif permettant d'instituer de nouveaux construits d'action collective"* (Crozier et Friedberg, 1977, p. 35). L'évaluation, par son apport rationnel, permettrait à l'organisation et à ses acteurs d'acquérir de nouvelles capacités.

Sans avoir pu approfondir l'apport de Crozier et Friedberg, ni utiliser la méthodologie de l'analyse stratégique des systèmes d'action, nous avons tenté d'intégrer une partie de leur réflexion en identifiant la position interne ou externe des acteurs des différentes composantes de l'évaluation par rapport à l'organisation qui mène l'action et par rapport au programme.

Outre la dimension implicite ou explicite délimitant huit types d'évaluation, on peut donc ajouter une dimension "interne ou externe" aux trois composantes de l'évaluation. Cette distinction supplémentaire s'applique donc aux processus d'évaluation, aux jugements de valeur et aux décisions.

1. Le processus d'évaluation est interne ou externe quand il est conduit par des personnes faisant partie ou non de l'équipe qui gère le programme.

C'est le sens courant du terme évaluateur interne ou évaluateur externe. Il indique la position de celui qui réalise l'acte d'évaluation. Qui peut conduire une évaluation ? Nevo (1986) fait part de quelques distinctions présentes dans la littérature.

La première concerne la position d'évaluateur externe ou interne selon l'appartenance ou non à l'équipe gérant un projet. Selon de nombreux auteurs, l'objectivité, la crédibilité de l'évaluateur interne seraient plus faibles, car moins indépendantes. Le degré d'indépendance de l'évaluateur et l'influence de cette position sur l'objectivité de son travail, notamment sur le jugement et la décision, est un enjeu souvent débattu. La formulation d'un processus d'évaluation exige la définition et le choix d'indicateurs de la réalité à décrire et donc une clarification de l'action et de ses objectifs. Il faudra aussi faire des choix dans la manière de recueillir l'information, de la traiter, de la communiquer. Un processus d'évaluation ne peut être totalement indépendant et objectif. Mais les choix peuvent être explicités. Notons qu'un évaluateur externe peut n'être pas indépendant (financièrement, institutionnellement).

La seconde distinction concerne la différence entre évaluateur professionnel et évaluateur amateur (Scriven, cité par Nevo, 1986). Il ne s'agit pas pour Scriven d'un jugement de valeur, mais d'une différence de formation et d'expérience. L'évaluation est, en effet, une véritable profession aux Etats-Unis, avec des formations universitaires spécifiques. Dans ce cadre, des auteurs ont défini les qualités du "bon" évaluateur : intégrité, objectivité, sensibilité au contexte social et aux rapports humains, mais aussi compétence dans les techniques de mesure. Des standards professionnels ont même été établis.

La plupart des auteurs nord-américains travaillent comme évaluateurs externes. En outre, même l'évaluation interne est conçue comme un travail spécifique nécessitant des compétences particulières au sein d'une équipe. Nous avons rarement trouvé dans la littérature américaine des références à l'adoption de plusieurs rôles par une même personne, par exemple : directeur de programme et évaluateur.

Pourtant, tous les acteurs d'un programme peuvent jouer un rôle dans l'évaluation. Il est plus pertinent d'envisager leur participation à l'évaluation sous la forme d'un continuum plutôt que d'oppositions. Ainsi, Sensi (1994) propose un continuum allant du statut de simple répondant à celui d'acteur dans l'auto-évaluation.

2. De la même manière, le jugement de valeur est interne ou externe en fonction des personnes qui émettent le jugement.

Barbier (1990, p. 238) pose de la manière suivante la question des rôles des différents acteurs impliqués dans le jugement de valeur. Il ne demande pas qui évalue, mais pour qui ou au nom de qui. "*Qui a le droit de porter un jugement de valeur sur la réalité évaluée ?*" Dans l'évaluation des actions, tous les acteurs ont le droit d'évaluer, mais le pouvoir d'évaluer n'est pas réparti de la même manière entre tous. Les rapports au droit d'évaluer sont déterminés par les conditions de détention du référent. "*Qui détient les objectifs qui servent à juger ?*"

Il faut donc nous semble-t-il ne pas se contenter d'identifier l'acteur qui procède à l'évaluation, mais aussi celui qui construit et porte le jugement sur le matériau recueilli. Mais la tâche est complexe parce que l'acte de juger n'est pas identifiable simplement au fait d'émettre un jugement, mais bien avant, à la détermination des critères ou des normes de jugement. C'est là, souvent, que le pouvoir de juger passe de l'acteur de terrain à l'évaluateur spécialisé.

L'identification du processus de description et de détermination des critères sera essentielle pour clarifier la production interne ou externe d'un jugement de valeur explicite. C'est pour Nevo (1986) la tâche la plus difficile de l'évaluation : choisir les critères pour juger la valeur d'un objet.

La position de celui qui détermine ces critères est déterminante. Ceux qui collectent des informations pour des décideurs ne se posent pas la question ou du moins elle reste implicite. Ceux qui mesurent si les objectifs fixés ont été atteints, n'ont pas beaucoup de difficulté non plus. Les critères proviennent d'un cadre fixé au préalable.

Quels sont les critères à utiliser pour juger de la valeur d'un objet ? D'autres se sont donné comme base de sélection des critères de jugement : les besoins de la société ou des idéaux (House), des besoins indiscutables découlant d'une situation donnée (Scriven), les souhaits de leurs clients (Patton), des normes relevant de professionnels (Eisner).

L'extériorité ou non du jugement de valeur soulève aussi la question du contrôle de l'action et des personnes qui la mènent. Sensi (1994) relève que cette extériorité introduit une dimension de menace qui peut provoquer chez les acteurs des réactions de type stratégique voire d'ordre affectif.

L'extériorité doit aussi être mise en relation avec la fonction de justification, le devoir de rendre compte (*accountability*) fixés à l'évaluation par différents protagonistes du programme. Le droit de contrôler, de juger est pour un auteur comme Mac Donald (cité par Sensi, 1994) un droit démocratique. House (1976) ne conçoit pas d'évaluation sans liberté de choisir. Il associe à cette liberté une société marchande, individualiste, rationnelle et compétitive. C'est-à-dire toutes les caractéristiques du Libéralisme prôné par John Stuart Mill.

C'est parce que le jugement est explicite, socialisé, mais aussi externe qu'il peut servir au contrôle ou à la justification. Si la norme est externe, standardisée, si le juge est extérieur au programme, la valeur sera différente. L'enjeu sera différent quel que soit l'objet évalué, processus ou produit, assurance de qualité ou évaluation d'efficacité.

La dimension interne ou externe du jugement de valeur peut aussi être appliquée au processus d'évaluation, au travail de celui qui conduit l'évaluation.

Qui évalue l'évaluation ? Quelles sont les normes de jugement de l'évaluation ?

Les auteurs américains sont peu explicites sur ce point, mais l'existence des "Standards" indique que la communauté des évaluateurs est la première concernée. Boruch et Cordray (cité par Stufflebeam, 1990) ont analysé six propositions de normes professionnelles de l'évaluation et concluent qu'elles sont très largement redondantes. La plus élaborée et celle qui fait le plus largement le consensus a été publiée par le *Joint Committee on Standards Educational Evaluation* en 1981 (Stufflebeam, 1990).

Pour Patton, l'évaluateur doit fournir des informations pertinentes pour des utilisateurs clairement identifiés. Le contrôle est exercé par les clients de l'évaluation.

La question du rapport final éclaire la position des invités de Alkin (1990) lors d'un débat. Pour Patton, le rapport final n'est pas une nécessité et se négocie. D'autres moyens d'information sont plus profitables et il n'est pas possible de répondre simultanément à tous les publics. Les tenants de la nécessité d'une audience par un public large refusent cette position et tiennent le rapport final pour indispensable. C'est la responsabilité de l'évaluateur et cela n'augmente pas les coûts. Un même rapport peut avoir plusieurs usages. Conner (cité par Alkin, 1990) insiste sur la nécessité du rapport final comme moyen de justifier le travail de l'évaluateur et de vérifier l'intégrité de celui-ci grâce à la revue par les pairs. Alkin s'étonne de ce que des personnes extérieures au contexte et ignorantes de l'impact, de l'utilisation de l'étude, puissent être sollicitées pour un examen critique, voire même pour la définition d'une bonne évaluation. Cette position revient à privilégier les aspects techniques et formels au détriment de la pertinence. Le débat entre positivistes et phénoménologistes se pose ici aussi.

Mais, le pouvoir de juger est aussi lié au pouvoir de décider et de conduire l'action, de l'intérieur ou de l'extérieur.

3. L'utilisation de l'évaluation est interne quand les décisions sont prises par les acteurs du programme sur des éléments de la conduite de l'action. Elle est externe quand ces décisions sont prises par des personnes extérieures.

A qui sert une évaluation ? Nevo (1986) reflète les positions de différents auteurs sur les utilisateurs de l'évaluation.

Scriven fait du "consommateur", du bénéficiaire du programme, le principal utilisateur de l'évaluation. Cronbach s'adresse à la communauté politique au sens large. Le public et la communauté scientifique sont aussi l'audience la plus importante, si ce n'est la plus directe, pour des auteurs comme Weiss ou pour les tenants de l'approche "scientifique" (*knowledge oriented evaluator*).

Stufflebeam envisage les gestionnaires des programmes et les responsables administratifs (bailleurs de fonds) comme les premiers interlocuteurs. Patton parle des clients directs de l'évaluation qui peuvent être les commanditaires, les responsables du programme, ceux qui y travaillent. Le premier travail de l'évaluateur est, pour Patton, l'identification de ces interlocuteurs.

Nevo (1986) retient de son analyse de la littérature les trois principes suivants : une évaluation peut avoir plus d'une audience ; différentes audiences peuvent avoir différents besoins d'évaluation ; ces audiences et ces besoins doivent être spécifiés aux premiers stades de l'évaluation. Ils vont influencer tout le processus d'évaluation : dans le choix des informations, leur recueil, le niveau d'analyse, le compte rendu et la diffusion des conclusions.

Sensi (1994) catégorise les utilisateurs en trois groupes : les SUPRA (décideurs politiques et administratifs), les INTRA (gestionnaires et participants) et les EXTRA (grand public, communauté scientifique, c'est-à-dire les utilisateurs secondaires). Nous distinguerons simplement des utilisateurs internes ou externes.

Quel lien entre la position des différents acteurs ?

L'application de l'axe interne - externe aux différentes composantes de l'évaluation risque d'entraîner une vision fragmentée du rôle des acteurs. En effet, un ou plusieurs acteurs peuvent intervenir sur un ou plusieurs éléments. De plus ces interventions peuvent être plus ou moins éloignées dans le temps.

Ces difficultés sont illustrées par le débat entre Alkin (1990) et ses invités sur l'interaction entre l'évaluateur (qui conduit l'évaluation) et les différentes catégories d'utilisateurs.

La position de Patton, pour qui l'évaluation doit répondre "aux utilisateurs désignés pour des usages désignés" (*intended user for intended use*), est au centre de la discussion. Selon son expression, l'évaluateur doit être interactif / adaptatif / réactif.

Cette position est critiquée assez durement par Weiss et House. La première, conformément à sa vision de chercheuse (*knowledge oriented evaluator*) se sent responsable d'explorer non seulement les sujets souhaités par l'utilisateur mais plus largement tous les aspects d'un programme, notamment en s'appuyant sur la littérature. Sa responsabilité scientifique va au-delà d'un besoin immédiat. Le second invoque une position éthique. Il est immoral de ne tenir compte que des intérêts de ses clients. Il prône une évaluation guidée par la justice sociale (*social justice-oriented evaluator*).

Pour Patton (*client-oriented evaluator*), la responsabilité de l'évaluateur se limite à son public direct et pas aux usages secondaires de l'évaluation. Ceux-ci ne sont pas un objectif important et ne sont pris en compte de façon active qu'à la demande des clients. Des considérations pratiques de ressources limitent de toute façon le champ des possibilités. Les audiences sont multiples, mais elles sont atteintes volontairement.

Alkin détaille les types d'interactions possibles dans la formulation des questions d'évaluation. Des différences dans les pratiques d'évaluation sont perceptibles. D'une part, aider réellement le client à découvrir ses propres questions, d'autre part guider le client vers les questions les plus pertinentes en relation avec les connaissances existantes. Le rôle des théories externes est discuté ; trois positions sont dégagées : choisir une théorie en fonction d'une situation et de sa propre expérience sans interaction avec les utilisateurs, aider le client à poser les bonnes questions en s'aidant des théories, proposer au client de revoir ses questions en fonction d'autres théories. La connaissance pratique du client est une théorie parmi d'autres, qui est analysée et testée au besoin et en accord avec celui-ci. Les connaissances théoriques doivent être utilisées pour explorer les variables critiques.

Le thème du praticien - évaluateur ouvre le débat sur d'autres questions.

Deux questions se posent : le praticien peut-il faire des évaluations et utilise-t-il plus les résultats dans ces conditions ? Des expériences de ce type ont été réalisées par le NIE, *National Institute of Education*, l'agence fédérale américaine qui finance les programmes novateurs en éducation et leur évaluation. Elle a développé au milieu des années septante le concept d'évaluation "par ceux qui tiennent les enjeux" (*stakeholders*) pour répondre aux nombreuses critiques formulées envers les pratiques d'évaluation et leur manque de résultats.

Les participants s'accordent pour dire que le praticien est capable d'évaluer, mais uniquement pour modifier ou améliorer le programme, pas pour décider de sa survie. Dans ce contexte, l'évaluation est directement utilisée dans la situation même de la pratique. Un conflit existe entre la capacité de gérer et de justifier simultanément. Ce seraient donc deux fonctions contradictoires et non complémentaires.

Ce bref compte rendu montre à la fois la complexité des rapports entre acteurs de l'évaluation, mais aussi l'importance conférée à la position externe de l'utilisateur par ces évaluateurs professionnels. La conduite d'une action au niveau interne de la gestion quotidienne d'un programme est distincte de sa conduite externe par un décideur politique ou administratif. Cette distinction mérite d'être relevée et discutée.

Spécificité et non spécificité !

En cours d'analyse, il est apparu important d'apporter une précision supplémentaire à la dimension interne - externe en notant le caractère spécifique ou non spécifique des composantes explicites externes.

Le processus, le jugement ou la décision peuvent ou non être extérieurs au promoteur et au programme, ils peuvent ou non être établis en rapport étroit avec la situation concernée. Cette information supplémentaire ne s'applique qu'aux composantes déjà explicitées et externes. On ne peut la préciser qu'après la définition du type d'évaluation. Elle renforce le critère d'extériorité des composantes de l'évaluation.

Le promoteur peut ainsi faire référence aux résultats d'une recherche (processus), à un expert (jugement) ou à un texte proposant des priorités comme une charte ou les imposant comme un traité (décision). Il nous semble nécessaire d'identifier si l'évaluation utilisée porte précisément sur la situation décrite par le projet comme son "terrain" d'action ou si elle n'a qu'une portée générale. C'est en quelque sorte une distinction interne - externe appliquée à la portée de l'argument.

Cette distinction permet de faire la différence entre la mise en œuvre d'une évaluation de besoins spécifiques menée par un évaluateur externe préalablement à un projet précis et la référence à la littérature internationale.

La spécificité est un indice de la manière dont les promoteurs utilisent l'évaluation dans le cadre de la formulation des objectifs ou de la planification. Nous préciserons ce cadre dans la partie de ce chapitre consacrée à l'évaluation de programmes.

II.C. LA DIMENSION TECHNIQUE OU L'AXE INFORMEL - FORMEL

La dimension technique est le troisième axe de différenciation des pratiques d'évaluation que nous avons retenu. Elle se caractérise par la formalisation ou non des composantes qui ont été préalablement explicitées.

La dimension méthodologique, technique de l'évaluation reste centrale dans les discours et les débats sur l'évaluation et mérite donc d'être prise en compte dans un modèle global d'identification des pratiques. Elle fait partie des trois dimensions retenues par Alkin et Ellett (1990) pour catégoriser les "modèles" d'évaluation. Elle correspond à la dimension pratique définie par Sensi (1994). Williams (1989) identifie aussi l'axe méthodologique comme un des principaux axes de différenciation entre les théories et les pratiques d'évaluation des principaux auteurs américains dans ce domaine.

La dimension méthodologique s'applique le plus souvent à la composante du processus d'évaluation. De nombreux auteurs en font une analyse particulièrement détaillée. L'évaluation dite "scientifique" ne se définit d'ailleurs souvent chez ses promoteurs que par cette dimension, elle est assimilée au paradigme expérimental. Cependant, dans sa recherche d'un savoir sur l'action, l'évaluateur utilise les méthodes de la science. Ce faisant, son intervention dans l'évaluation prend une signification comme pratique sociale, dans la mesure où elle joue un rôle particulier en relation avec la rationalité et l'apprentissage collectif.

Pour généraliser cette dimension méthodologique et pouvoir la mettre en œuvre sur les trois composantes, nous la définissons par un axe formalisation - non formalisation. La formalisation de la pratique concerne donc le processus d'évaluation, la formulation du jugement de valeur et l'utilisation de l'évaluation.

1. La formalisation du processus d'évaluation porte sur plusieurs points correspondant aux différentes étapes du processus.

Quel est le processus de réalisation d'une évaluation ?

Les étapes distinguées dans le processus d'évaluation varient selon les auteurs. On peut tenter une synthèse en prenant comme fil conducteur les trois étapes fondamentales proposées par Stufflebeam et ses collègues (1971, traduction en français en 1980).

a) Délimiter les informations à recueillir.

Nevo (1986) présente plusieurs démarches. Il s'agit d'identifier les buts et de les transformer en objectifs mesurables chez les tenants de l'approche "objectifs - performances". Provus, par exemple, parle de clarifier le schéma du programme. Parlett et Hamilton passent aussi par une phase d'identification du contexte de l'évaluation, comme Patton qui demande d'identifier de manière personnelle les utilisateurs et leurs besoins.

Cette phase d'analyse préalable se retrouve partout, mais selon les options des auteurs, les interactions avec les utilisateurs y sont plus ou moins accentuées.

Barbier (1990, pp. 65 à 72) qui analyse l'évaluation comme un procès de travail qualifie cette étape de matériau ou de référent de l'évaluation.

Le référent est "ce à partir de quoi" un jugement de valeur est porté. C'est une représentation sur les faits, qui appelle une description et une mesure, qui fonde l'idéologie de l'objectivité et le développement d'un appareillage statistique.

Le référent est constitué, produit, ce n'est pas un fait naturel. La constitution du référent comporte deux étapes : le choix d'indicateurs de la réalité à décrire et le recueil des informations. Cette deuxième étape correspond à la deuxième étape proposée par Stufflebeam.

Tout d'abord, il faut choisir et désigner par des indices les caractéristiques qui seraient représentatives. Cette construction pose trois problèmes : le degré et le moment de l'explicitation, la nature de l'indicateur qui doit être représentatif et accessible, les conditions de détermination. Cette détermination des indicateurs de la réalité est liée à la définition des objectifs de l'action. Quand il s'agit d'évaluer des actions, la définition du référent demande au préalable une clarification de l'action, une théorie de l'action, puis une clarification du référent, c'est-à-dire des objectifs, et enfin des rôles joués. Ainsi, bien que cette étape soit d'ordre méthodologique, elle n'échappe cependant pas aux rapports entre acteurs.

b) Obtenir les informations souhaitées par une démarche formelle et des procédures d'analyse.

Tous les modèles passent par cette phase, mais en utilisant des méthodes différentes : des démarches qualitatives, des instruments quantitatifs, des outils d'analyse statistique, fixés au départ ou évoluant au cours du processus. Le moment du recueil d'information est défini par le type d'objet analysé avant, pendant ou après l'action.

Pour Barbier, le processus de production de l'information est un processus social complet avec des instruments (recueil de données existantes, observation, questionnement, expérimentation) ; des rapports sociaux entre observateur et observé ; une forme de représentation et de traitement de l'information.

c) Fournir les informations.

La position des auteurs varie sur ce point en fonction de leur position vis-à-vis du jugement de valeur : faire des recommandations, porter un jugement ou se limiter à fournir les informations. Tous les modèles prévoient cependant une phase de communication. Tous se situent en effet dans la perspective d'un intervenant externe au programme.

d) L'enjeu social de la formalisation du processus d'évaluation.

L'analyse des positions à propos de ces trois étapes du processus d'évaluation est traversée par le débat sur les méthodes de recherche à utiliser en évaluation. Ce débat permet de situer l'enjeu social de la formalisation et ce faisant de justifier le choix de la formalisation comme critère d'analyse de la dimension méthodologique de préférence à une typologie des méthodes.

Le débat entre méthodologie qualitative et quantitative existe toujours et a ses représentants déclarés. Cependant, l'éclectisme semble maintenant de mise, qui suggère un vaste choix de méthodes en fonction des besoins particuliers d'une évaluation. Mais cette position connaît aussi ses détracteurs.

Selon l'étude de Williams (1989), ce débat entre le quantitatif et le qualitatif, entre une vision positiviste, explicative ou naturaliste, compréhensive est le point de différenciation le plus important entre les auteurs. Mais c'est en fait plus un choix idéologique que technique, qui renvoie à la position de l'évaluateur dans le processus de l'évaluation et de l'action.

Bryman (1984) pose judicieusement la question sur le plan des sciences sociales. Il qualifie d'artificielle la distinction entre le quantitatif et le qualitatif sur le plan des méthodes, de la technique. Il y a un lien entre le problème de recherche ou d'évaluation à résoudre et le choix de telle ou telle méthode. Le pluralisme méthodologique s'impose donc de lui-même.

Par contre le débat épistémologique semble plus fondamental et correspondre à des manières différentes d'envisager et d'interpréter les phénomènes. Ce double constat peut sembler contradictoire. Bryman (1984) fait cependant remarquer que le lien entre une position épistémologique et le choix d'une technique est loin d'être clair et symétrique. Il cite une étude de Snizek qui montre que l'analyse des articles de recherche ne permet pas d'extrapoler les hypothèses épistémologiques des auteurs à partir du choix des techniques. La préférence pour un style de recherche semble plutôt liée à la formation et au contexte institutionnel.

L'identification de la technique de recueil d'information ne semble donc pas un indicateur précis. Le choix de la méthode peut être fonction du problème à résoudre, mais aussi de choix épistémologiques ou encore de contraintes institutionnelles.

Une explication de l'utilisation prioritaire, en évaluation, du modèle de la recherche expérimentale est proposée par Alkin et ses invités au cours de leur débat (Alkin, 1990). Ils l'attribuent à la légitimation professionnelle. L'évaluateur joue la sécurité pour ne pas être critiqué, il recourt donc au modèle de recherche qui a la tradition la plus forte. On pourrait y ajouter la nécessité d'une reconnaissance académique et l'accès aux revues avec comité de lecture. Même si d'autres modèles sont reconnus, CIPP ou "évaluation répondante", beaucoup hésitent à les utiliser ou à déclarer qu'ils les utilisent. Patton a remarqué que son modèle a apporté une légitimation à beaucoup d'évaluateurs plus centrés sur les situations locales. Une des études de Alkin montre que les évaluateurs font très rarement référence à des modèles, bien qu'ils suivent en fait certaines options reconnaissables, issues de leur connaissance pratique ou de leur formation antérieure.

De ces considérations sur le débat méthodologique, nous retenons que la présence ou l'absence d'une formalisation du processus est plus significative pour différencier les pratiques que des caractéristiques méthodologiques telles que choix d'une méthode quantitative ou qualitative, par exemple.

Nous déterminerons donc qu'un processus est formel s'il présente des indicateurs (phase de délimitation), une méthode et des outils de recueil de l'information (phase de recueil) et une phase de communication des informations.

2. Le jugement de valeur est considéré comme formalisé quand les critères, les normes, les valeurs sur lesquels porte le jugement sont présentés.

Le jugement n'est plus seulement énoncé (explicitation) et identifié comme celui d'un acteur interne ou externe au programme. Il est aussi justifié par rapport à des valeurs préalablement définies. Cette formalisation permet la communication voire la contradiction.

Barbier (1990, pp. 72 à 78) précise que cette étape nécessite de définir le moyen, le référent de l'évaluation. Le référent est "ce par rapport à quoi" le jugement de valeur est porté. Ce sont les normes de jugement, les critères, les points de repères. Le référent est de l'ordre des objectifs, des valeurs. Il joue un rôle d'instrument dans la production du jugement de valeur.

Les critères et les normes n'existent pas à l'état naturel. Le plus souvent, ils sont subjectifs, implicites. Il y a d'abord identification des objectifs poursuivis. Ensuite vient la spécification des normes ou critères. Les critères doivent permettre une comparaison par un processus de rationalisation et de standardisation. C'est cette construction du référent qui apporte la formalisation du jugement de valeur.

3. La formalisation de l'utilisation de l'évaluation.

Une décision peut être énoncée, attribuée à un acteur. Elle peut aussi être formalisée. Les éléments de cette réflexion sont fournis par les recherches sur l'utilisation de l'évaluation, que nous détaillerons dans le troisième chapitre.

Leviton et Hughes (1981) identifient différents types d'utilisation, d'après une revue de littérature des recherches à ce propos.

Précisons tout d'abord que le processus de communication que nous avons identifié plus avant n'est pas une utilisation. La lecture d'un rapport d'évaluation n'est pas une "utilisation". En effet, l'utilisation implique un changement et ce changement prend principalement deux formes.

La première qualifiée d'instrumentale par ces deux auteurs est une utilisation directe, observable, le plus souvent immédiate. Par exemple, un programme se voit prolongé après une évaluation d'efficacité positive.

La deuxième, conceptuelle, n'est pas directement observable. Elle porte sur la manière de penser, de juger un programme. L'évaluation est une source d'information pour le décideur, sans utilisation directe immédiate.

Pour déterminer la formalisation de l'utilisation de l'évaluation, nous avons pris comme indicateur l'influence directe de la décision sur le programme et le moment de cette décision, immédiate ou différée.

III. LA GRILLE D'ANALYSE APEP

Tous les ingrédients nécessaires à la construction de l'instrument d'analyse des pratiques d'évaluation de programmes sont maintenant définis : les trois composantes de l'évaluation et les trois dimensions d'une pratique d'évaluation.

La combinaison de ces différents éléments a été réalisée en deux étapes. La première combine les trois composantes de l'évaluation avec l'axe implicite - explicite pour définir huit types d'évaluation. Cette première grille servira d'ailleurs de fil conducteur dans l'examen des pratiques théorisées (chapitres 2, 3 et 4).

La seconde croise ces huit types d'évaluation (et leurs trois composantes) avec les axes stratégique et technique (tous deux bipolaires), pour obtenir un modèle à trois voire à quatre dimensions selon les cas.

III.A. LES HUIT TYPES D'EVALUATION

Une description plus détaillée de ces huit types d'évaluation déjà présentés ci-dessus (en II.A.) permet d'en préciser le contenu. De nombreux exemples seront en outre proposés dans le chapitre 2.

T0 : L'évaluation implicite

Processus	Jugement de valeur	Décision	Explicite
			Implicite

Les trois composantes évaluatives sont implicites. Aucun élément ne trouve spontanément une énonciation "socialisée" Le programme se déroule sans manifestation extérieure sur sa conduite.

Toute l'évaluation se passe au niveau interne de l'action, qui s'auto-régule, ou au niveau de personnes extérieures qui jugent la valeur du programme ou prennent des décisions, mais sans les communiquer. Un observateur extérieur ne peut donc identifier cette forme d'évaluation qu'en interrogeant les acteurs du programme. Ce type d'évaluation ne fera donc pas l'objet de l'analyse des documents-projets puisqu'il n'y a pas d'énonciation, mais il sera abordé lors des entretiens avec certains promoteurs.

T1 : L'évaluation "décision" (décision explicite)

Processus	Jugement de valeur	Décision	Explicite
			Implicite

La décision est explicite, socialisée. Le processus et le jugement de valeur conduisant à cette décision restent cependant cachés. Le décideur occupe une position sociale reconnue permettant la décision au niveau interne ou externe. Cette décision peut être immédiate, toucher directement au programme ou être différée.

La gestion d'un programme implique aussi bien le responsable de sa conduite que le commanditaire qui le finance. Ce gestionnaire peut, bien sûr, utiliser des types d'évaluation dont les processus et les jugements sont explicités. Mais, si certaines décisions sont explicitées, le processus qui y conduit n'est pas rappelé. C'est le cas par exemple de l'octroi d'une subvention, de la formulation d'une priorité de santé publique ou du choix d'un outil pédagogique avant une séance d'animation. Dans le cadre de cette typologie, nous avons qualifié ces décisions d'évaluation de gestion. Il serait plus précis de dire "évaluation implicite de gestion".

T2 : L'évaluation d'expertise (jugement explicite)

Jugement de valeur		<i>Explicite</i> <hr style="border-top: 1px dashed black;"/> <i>Implicite</i>
Processus	Décision	

Il y a un jugement de valeur explicite, socialisé, mais le processus aboutissant au jugement reste implicite de même que la décision qui découle du jugement.

Le jugement de valeur est souvent confié à une personne extérieure à l'action, mais il peut être émis par une personne interne reconnue pour cette fonction. Cette évaluation est donc organisée ou réalisée par des personnes socialement déléguées pour le faire. Le jugement peut aussi être formalisé en référence à des normes établies.

T3 : L'évaluation formelle (processus explicite)

Processus	<hr style="border-top: 1px dashed black;"/> Jugement de valeur Décision		<i>Explicite</i> <hr style="border-top: 1px dashed black;"/> <i>Implicite</i>

Seul le processus est explicite, énoncé.

Le modèle le plus fréquent sera probablement celui de l'évaluation professionnalisée menée par un évaluateur "objectif", externe à l'action, qui va décrire et détailler son processus de travail (formalisation). Mais d'autres pratiques, internes, formalisées ou non, pourront exister. Les jugements de valeur et les décisions ne sont pas explicités. La fonction principale est de connaître et comprendre un programme.

T4 : L'évaluation de contrôle (jugement et décision explicites)

Jugement de valeur Décision		<i>Explicite</i> <hr style="border-top: 1px dashed black;"/> <i>Implicite</i>
Processus		

Il y a une décision et un jugement de valeur, le plus souvent à partir de normes fixées (formalisées), mais le processus conduisant au jugement reste implicite. C'est l'expression du jugement de valeur qui différencie cette évaluation de l'évaluation "décision" (type I). Les personnes qui jugent et décident peuvent être les mêmes ou non. Elles peuvent être internes ou externes à l'action. Elles sont socialement habilitées pour juger et décider.

T5 : L'évaluation orientée vers les conclusions (processus et jugement explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

Les deux premières composantes sont explicites. Le processus d'évaluation et le jugement de valeur sont énoncés. La fonction de ce type d'évaluation est principalement la production d'un jugement de valeur. C'est l'évaluation de la valeur d'un programme, de son efficacité. La décision est absente ou elle reste implicite.

Ce type d'évaluation est le plus souvent réalisé par un évaluateur externe professionnel, qui formalise un processus et un jugement. La conduite de l'action, la décision appartient alors au directeur du programme ou à une personne extérieure et ne sera pas explicitée. La finalité principale est la justification (*accountability*). D'autres pratiques internes et / ou moins formalisées devraient exister.

T6 : L'évaluation "processus - décision" (processus et décision explicites)

Processus	Décision	<i>Explicite</i>
Jugement de valeur		<i>Implicite</i>

Le jugement de valeur est le seul à être implicite. Il y a un processus d'évaluation déclaré et mené de manière explicite pour aboutir à des décisions identifiables, énoncées. Nous pourrions identifier différentes formes d'évaluation orientée vers la décision selon la dimension interne - externe du processus et de la décision.

Le caractère immédiat ou différé de la décision (axe informel - formel) est aussi une caractéristique importante pour identifier ces pratiques à visée "formative" qui cherchent à diagnostiquer rapidement les difficultés afin d'y remédier.

T7 : L'évaluation explicite (processus, jugement et décision explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

Les trois composantes sont explicites. Tout est décrit, expliqué, communiqué à l'extérieur du programme. Il y a un processus d'évaluation identifiable. Il y a un jugement de valeur et des décisions sur la conduite de l'action qui sont énoncés. Chaque composante pourra aussi être décrite sur les axes interne - externe et informel - formel.

III.B. LES 124 MODALITES D'EVALUATION DE LA GRILLE D'ANALYSE

12 modalités		T1 Evaluation "décision"			T2 Evaluation d'expertise			T3 Evaluation formelle		
		P	J	D	P	J	D	P	J	D
Stratégique	Sociale									
	Technique									
Interne Informel				Dii		Jii		Pii		
Interne Formel				Dif		Jif		Pif		
Externe Informel				Dei		Jej		Pei		
Externe Formel				Def		Jef		Pef		

48 modalités			T4 Evaluation contrôle			T5 Evaluation conclusion			T6 Evaluation "processus - décision"		
			P	J	D	P	J	D	P	J	D
Stratégique	Sociale										
		Technique									
Interne Informel	x	ii		Jii	Dii	Pii	Jii		Pii		Dii
		if			Dif		Jif				Dif
		ei			Dei		Jej				Dei
		ef			Def		Jef				Def
Interne Formel	x	ii		Jif	Dii	Pif	Jii		Pif		Dii
		if			Dif		Jif				Dif
		ei			Dei		Jej				Dei
		ef			Def		Jef				Def
Externe Informel	x	ii		Jej	Dii	Pei	Jii		Pei		Dii
		if			Dif		Jif				Dif
		ei			Dei		Jej				Dei
		ef			Def		Jef				Def
Externe Formel	x	ii		Jef	Dii	Pef	Jii		Pef		Dii
		if			Dif		Jif				Dif
		ei			Dei		Jej				Dei
		ef			Def		Jef				Def

64 modalités	T7 Evaluation explicite									
Sociale	P et J et D explicites									
Stratégique / Technique	Interne Informel		Interne Formel		Externe Informel		Externe Formel			
	Pii x		Pif x		Pei x		Pef x			
Interne Informel	x	ii	Jii	Dii	Jii	Dii	Jii	Dii	Jii	Dii
		if		Dif		Dif		Dif		Dif
		ei		Dei		Dei		Dei		Dei
		ef		Def		Def		Def		Def
Interne Formel	x	ii	Jif	Dii	Jif	Dii	Jif	Dii	Jif	Dii
		if		Dif		Dif		Dif		Dif
		ei		Dei		Dei		Dei		Dei
		ef		Def		Def		Def		Def
Externe Informel	x	ii	Jei	Dii	Jei	Dii	Jei	Dii	Jei	Dii
		if		Dif		Dif		Dif		Dif
		ei		Dei		Dei		Dei		Dei
		ef		Def		Def		Def		Def
Externe Formel	x	ii	Jef	Dii	Jef	Dii	Jef	Dii	Jef	Dii
		if		Dif		Dif		Dif		Dif
		ei		Dei		Dei		Dei		Dei
		ef		Def		Def		Def		Def

La combinaison des trois composantes et des trois dimensions prédit donc l'existence de 124 modalités théoriques différentes. L'analyse des modèles existants et l'observation des pratiques utilisées dans les programmes devront permettre de faire apparaître des formes plus ou moins fréquentes et de les mettre en relation avec des conditions particulières d'apparition qui les expliqueraient.

Pour décrire ces pratiques d'évaluation dans les programmes d'éducation pour la santé, nous avons notamment formalisé la structure et le contenu de ces programmes. Chaque pratique identifiée avec la grille d'analyse des pratiques d'évaluation de programmes sera ainsi mise en relation avec la structure et le contenu du programme.

IV. LA DESCRIPTION DES PROGRAMMES

"*Une civilisation des projets ?*", c'est sur ce titre que commence l'ouvrage de Barbier publié en 1991 (Elaboration de projets d'action et planification). Ce développement de la gestion des activités sociales par les projets est en effet une caractéristique importante de l'évolution actuelle des politiques menées en ce domaine et le secteur de la promotion de la santé en Communauté française de Belgique n'y échappe pas. Le récent décret organisant la promotion de la santé, comme celui portant sur l'école, en sont des exemples.

On pourrait s'interroger sur cette évolution, y voir, comme Barbier (1991), le signe d'un nouveau mode d'organisation du travail permettant plus d'autonomie, de cohérence, de développement aux acteurs. Mais on peut y voir aussi un nouveau mode de contrôle social où la responsabilité des choix est déplacée du politique vers les acteurs sociaux tout en limitant leurs moyens. En effet, souvent, un projet s'inscrit dans un laps de temps plus court qu'un service.

La place de l'évaluation dans cette évolution est aussi croissante. Elle est intimement liée aux concepts de projet et de programme. C'est ce lien que nous allons présenter en nous appuyant sur le livre de Barbier (1991), mais aussi sur l'ouvrage de Stufflebeam et de ses collègues publié une quinzaine d'années plus tôt (1971, 1980 pour la version en français). Nous établirons également un parallèle avec notre travail de soutien aux projets d'éducation pour la santé en Communauté française depuis 1982. Une brochure (Demarteau, 1992b) présente une méthodologie de construction de programme en éducation pour la santé très proche des deux modèles précédents.

Nous utiliserons la dénomination "programme" plutôt que celle de "projet" pour des raisons théoriques et pratiques. Le projet, selon Barbier (1991) est une représentation, "*une image anticipatrice d'un processus de transformation du réel.*" (p. 84) Le programme, selon le même auteur, est un document écrit, un "document-projet", une énonciation sociale, un acte délibéré, socialement organisé. "*Tout projet mérite donc d'être analysé au niveau du langage qu'il exprime ; (...) il dit quelque chose sur l'acteur qui le produit et son environnement.*" (p. 265)

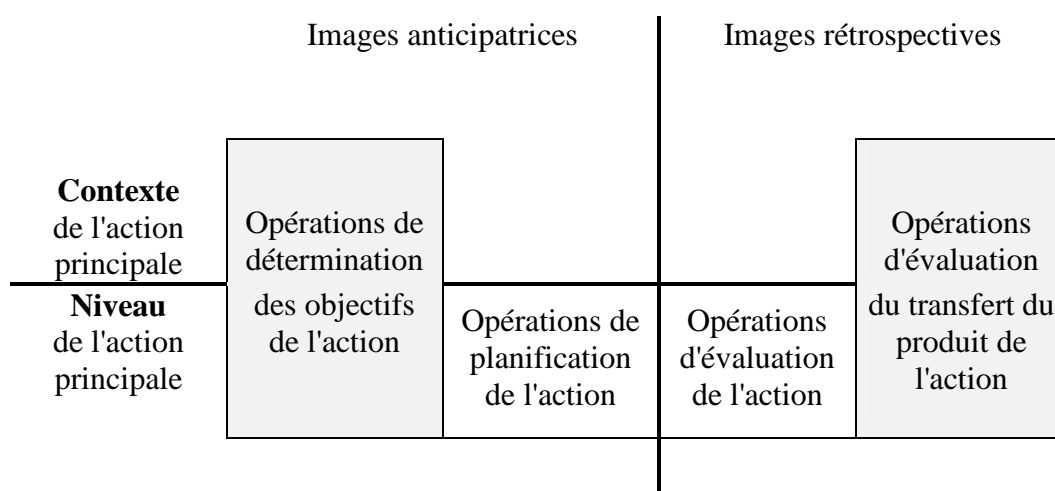
L'analyse des programmes écrits a été privilégiée dans cette étude en raison de leur accessibilité. Dans la mesure où ces programmes écrits sont considérés comme des productions sociales, ils sont un support adéquat pour identifier des pratiques d'évaluation explicitées.

IV.A. LA STRUCTURE D'UN PROGRAMME

1. Le modèle de Barbier : le procès de conduite des actions.

En présentant son analyse de la démarche de planification, Barbier précise que celle-ci n'existe pas indépendamment d'autres représentations. Cette chaîne de représentations, il la qualifie de "*procès de conduite d'une action*" (pp. 88 à 94). Il met ainsi en évidence le lien avec ces deux autres ouvrages sur l'analyse des besoins en formation (1986) et l'évaluation en formation (1990). C'est ce procès qui permet d'identifier les grandes étapes de la conduite d'une action, de la constitution d'un projet, d'un programme.

Organisation d'ensemble du procès de conduite d'une action (Barbier, 1991, p. 92)



La première étape est la détermination des objectifs finaux d'une action, la représentation de l'état final. Les opérations qui y mènent ont pour noms : analyse des besoins, analyse de la demande, analyse diagnostique, analyse stratégique, analyse de la situation, etc.

La deuxième étape comporte les opérations qui permettent de définir le plan d'action, la représentation du processus de transformation : l'inventaire des moyens d'action, l'évaluation des intrants, le choix des méthodes, la définition des rôles, etc.

Ces deux étapes sont des anticipations.

La troisième a pour fonction de produire un jugement de valeur sur l'action principale en référence aux objectifs et au plan : évaluation de processus, évaluation interne, régulation, tableau de bord, évaluation de résultats, etc.

La quatrième produit un jugement de valeur sur la mise en œuvre plus large en référence à la situation qui a suscité l'élaboration de l'action principale : évaluation externe, évaluation de l'impact, évaluation des effets, etc.

Ces deux dernières étapes sont des vues rétrospectives sur l'action et son contexte.

Ces opérations sont cohérentes les unes vis-à-vis des autres, et cycliques. Elles sont dépendantes, mais pas linéaires. Il y a une continuité entre l'émergence d'un problème et l'évaluation des effets de l'action mise en place pour le résoudre. C'est cette continuité et la place qu'y jouent différentes pratiques d'évaluation que nous étudions.

Les quatre étapes présentées par Barbier semblent bien caractériser une démarche de conduite d'une action, bien que nous ne partageons pas entièrement son analyse sur les opérations à faire figurer dans les catégories créées, notamment pour les catégories qu'il appelle "évaluation".

2. Le modèle CIPP de Stufflebeam.

L'ouvrage de Stufflebeam (1980) permet de positionner autrement l'évaluation dans ce cycle de la conduite d'une action, d'une part, en assimilant les opérations de constitution des objectifs et du plan à des opérations d'évaluation et d'autre part, en distinguant l'évaluation rétrospective sur le plan de l'évaluation rétrospective sur les objectifs.

Il y a une étonnante similitude entre le modèle proposé par Barbier et celui proposé de nombreuses années auparavant par Stufflebeam. Le tableau suivant le met en évidence.

Les types de décisions (Stufflebeam, 1980, p. 101)

	Projeter	Réaliser
Les Fins	LES DECISIONS DE PLANIFICATION déterminer les objectifs	LES DECISIONS DE REVISION jauger les réalisations
Les Moyens	LES DECISIONS DE STRUCTURATION délimiter les procédures	LES DECISIONS DE MISE EN APPLICATION contrôler les procédures

On y retrouve la distinction entre l'anticipation et la rétrospection, entre les fins (objectifs) et les moyens (plan), même si le terme planification est pris dans un sens différent chez les deux auteurs.

C'est par rapport à cette classification des décisions que Stufflebeam va situer le modèle d'évaluation intitulé CIPP (Contexte, Intransigent, Processus, Produit).

C'est l'expérience de l'évaluation des projets ESEA (*Elementary and Secondary Education Act*) dans les écoles publiques, en 1965, qui conduit Stufflebeam à critiquer les méthodes existantes centrées sur l'approche tylerienne et les tests psychométriques (Stufflebeam et Shinkfield, 1985, pp. 151 à 207). Stufflebeam propose alors de redéfinir l'évaluation en la liant très clairement à la prise de décision. Il identifie différentes catégories de décisions qui correspondent à différents types d'évaluation. Les décisions d'implantation demandent une attention particulière aux processus tandis que les décisions de sélection, de subvention s'appuient sur les résultats, les produits. Sa réflexion s'élargit ensuite aux décisions de planification qui s'intéressent aux buts, aux objectifs et aux décisions de structuration basées sur l'analyse des moyens.

a) L'évaluation de contexte contribue à déterminer les objectifs.

L'évaluation de contexte définit l'environnement en cause (ses valeurs et ses buts). Elle décrit les conditions existantes et souhaitées relatives à cet environnement. Elle identifie la population cible et les besoins à combler, diagnostique les problèmes qui empêchent que ces besoins soient satisfaits et les occasions que l'on pourrait saisir. Elle juge si les objectifs proposés répondent suffisamment aux besoins mesurés ou si d'autres buts doivent être poursuivis. Elle fournit les bases pour décider des buts, des objectifs, des changements à effectuer et pour juger plus tard des résultats.

Ces informations peuvent être trouvées dans les textes officiels, mais aussi dans des discours, des revues de presse ou des sondages d'opinions, auprès d'experts ou de témoins. L'évaluation de contexte utilise aussi des enquêtes, des revues de documents, des auditions, des entretiens, des tests diagnostiques, la technique Delphi, etc.

Les décideurs clés sont les décideurs du système, en particulier au plan politique et administratif, mais aussi plus globalement les acteurs du système. C'est auprès de ceux-ci que l'évaluateur va délimiter, obtenir et fournir les informations utiles.

b) L'évaluation d'intrant aide à déterminer les structures des projets.

L'évaluation d'intrant inventorie et analyse les ressources matérielles et humaines disponibles, les solutions proposées et les stratégies alternatives au programme. Elle fournit les bases pour juger des capacités du système pour implanter les stratégies. Elle identifie et mesure les supports, les solutions, les budgets et les délais pour apprécier la pertinence, la faisabilité et l'économie des plans de procédure et des stratégies envisagées.

Cette évaluation se fait notamment par la confrontation à d'autres expériences, mais elle utilise aussi d'autres méthodes comme la recherche de littérature, des visites à des programmes exemplaires, des procédures d'examen contradictoire et des essais pilotes à petite échelle.

Le but est d'aider le client, le directeur de programme, à considérer des stratégies de programme alternatives en fonction du contexte et des besoins, des contraintes et des ressources.

c) L'évaluation de processus aide à contrôler les opérations.

L'évaluation de processus identifie les défauts dans les procédures planifiées. Elle fournit des informations spécifiques pour des décisions programmées. Elle enregistre et juge les activités au cours du programme. Elle assure un contrôle effectif du processus. Elle surveille les barrières potentielles aux activités et reste en alerte vis-à-vis des difficultés non prévues. Elle décrit le processus en cours en observant les activités de l'équipe du projet. Elle fournit les bases pour juger le déroulement du processus et assurer l'interprétation future des résultats.

Elle utilise l'observation, l'enregistrement de données, le journal de bord, toutes les techniques d'enquête, mais de façon brève et répétée.

Elle concerne l'implantation, le raffinement du plan et des procédures du programme, le respect de la planification, l'utilisation des ressources, l'atteinte des buts intermédiaires, l'analyse du rôle des participants, le suivi des coûts réels.

Pour Stufflebeam, c'est une tâche particulière et nécessaire qui doit être assumée par une personne particulière. Le travail quotidien ne fournit pas une information suffisante pour ce type de décision. Les interlocuteurs sont les gestionnaires et les acteurs du programme.

d) L'évaluation de produit aide à juger et à réviser les projets.

L'évaluation de produit collecte les descriptions et les jugements sur les résultats. Elle les met en relation avec le contexte et les objectifs, les intrants et le processus. Elle mesure, interprète et juge la réussite du programme.

Elle définit opérationnellement des critères de jugement, collecte des informations et réalise des analyses qualitatives et quantitatives. Elle présente clairement les effets enregistrés (attendus et inattendus, positifs et négatifs).

Elle permet des décisions sur la continuation, l'arrêt, la modification ou la relocalisation d'une activité.

3. Les catégories d'analyse de la structure d'un programme.

Avec son modèle, publié pour la première fois en 1971, Stufflebeam va le premier élargir et structurer l'évaluation de programmes. Il va clairement situer l'enjeu de l'évaluation en relation avec la conduite des programmes par l'aide à la décision. Il va envisager aussi bien le contexte général du programme que ses objectifs et ses moyens particuliers.

C'est ce modèle et celui de Barbier qui nous ont inspiré dans l'identification et la description des pratiques d'évaluation de programmes d'éducation pour la santé. Nous avons toutefois articulé les diverses étapes de la conduite de l'action en identifiant les décisions proprement dites selon le schéma suivant.

La structure d'un programme

PROGRAMME					
PROJET D'ACTION				PROJET D'EVALUATION	
EVALUATION DE CONTEXTE	OBJECTIFS	EVALUATION D'INTRANT	PLAN	EVALUATION DE PROCESSUS	EVALUATION DE PRODUIT
Opération	Représentation	Opération	Représentation	Représentation	Représentation
Fins		Moyens		Moyens	Fins

A partir de la structure d'un programme, présentée ci-dessus, nous avons tenté d'identifier les pratiques d'évaluation correspondant au modèle CIPP. Dans l'évaluation de contexte, nous situons tous les arguments permettant de déterminer les objectifs ou de les justifier. Dans l'évaluation d'intrant se placent les arguments qui servent à déterminer ou justifier les activités et les ressources à mettre en œuvre. Dans l'évaluation de processus et de produit sont compris les projets d'évaluation portant sur les moyens et sur les fins.

Il nous a fallu aussi considérer que l'ensemble du programme (document-projet) y compris l'évaluation de processus et de produit se situe dans une vision anticipatrice. Il y a projet d'évaluation comme il y a projet d'action. Ainsi les pratiques d'évaluation qui feront l'objet de l'analyse seront assimilables soit à des opérations (évaluation de contexte et d'intrant), soit à des représentations (objectifs, plans, évaluation de processus et de produit).

IV.B. LE CONTENU D'UN PROGRAMME

Outre la structure du programme, nous nous sommes attaché également à en décrire le contenu. Ce contenu est spécifique des programmes d'éducation pour la santé et de promotion de la santé.

Les catégories créées permettent de rendre compte de différents types d'objectifs et de plans, qui sont également les objets de l'évaluation de contexte, d'intrant, de processus et de produit, conformément à la structure de programme en miroir proposée dans le modèle ci-dessus.

1. Les objectifs et objets de l'évaluation de contexte et de produit.

Neuf catégories d'objectifs ont été identifiées au départ du modèle PRECEDE de Green (1980). Ce modèle de planification, créé spécifiquement pour l'éducation pour la santé, rencontre nos besoins.

1. Les objectifs socio-économiques.
2. Les objectifs sanitaires.
3. Les objectifs environnementaux.
4. Les objectifs comportementaux.
5. Les objectifs cognitifs.
6. Les objectifs affectifs.
7. Les objectifs psychomoteurs.
8. Les objectifs administratifs.
9. Les objectifs d'autonomie.

Comme dans le modèle PRECEDE, les catégories illustrent un diagnostic en entonnoir partant de l'évaluation d'une situation sociale globale (culturelle, économique), puis passant par l'évaluation de l'état sanitaire et de l'environnement matériel pour arriver à l'évaluation comportementale, éducative et finalement administrative (les structures). Nous nous sommes écarté du modèle PRECEDE en utilisant les trois catégories d'objectifs éducatifs classiques (cognitifs, affectifs et psychomoteurs) définies par Bloom (cité par De Landsheere et De Landsheere, 1975) plutôt que la distinction entre facteurs prédisposants, renforçants et facilitants, jugée peu opérationnelle.

La catégorie "objectifs administratifs" appelle une précision. Il ne s'agit pas de l'évaluation des ressources administratives nécessaires pour réaliser le programme, auquel cas elle figurerait dans l'évaluation d'intrant, mais d'une évaluation du contexte administratif qui influence le problème de santé à résoudre, par exemple la communication entre le médecin et son patient.

Nous avons, enfin, introduit une dernière catégorie en cours d'analyse pour rendre compte du concept d'autonomisation individuelle et collective que nous rencontrons dans certains projets.

2. Le plan et les objets de l'évaluation d'intrant et de processus.

Treize catégories ont été identifiées à partir de notre expérience et de l'analyse préalable de quelques programmes.

- | | |
|--|---|
| 1. Les actions socio-politiques. | 7. Les actions d'enquêtes. |
| 2. Les actions environnementales. | 8. Les actions de prévention sanitaire. |
| 3. Les actions communautaires. | 9. Les ressources humaines. |
| 4. Les actions éducatives. | 10. Les ressources matérielles. |
| 5. Les actions de communication sociale. | 11. Les ressources financières. |
| 6. Les actions de formation. | 12. Les ressources temporelles. |
| | 13. La gestion globale. |

Les huit premières catégories recouvrent les stratégies d'actions et les cinq autres les ressources à mobiliser.

L'action socio-politique concerne l'utilisation des structures sociales organisées et préexistantes. L'éducation pour la santé atteint la population à travers la mobilisation de structures sociales intermédiaires.

L'action environnementale s'attache à modifier les conditions matérielles, par exemple l'offre alimentaire.

L'action communautaire met l'accent sur la participation active de la communauté.

L'action éducative vise l'apprentissage, y compris par les moyens audiovisuels.

L'action de communication sociale exploite les médias pour sensibiliser et informer.

L'action de formation s'adresse à des relais vers la population visée.

L'action d'enquête met l'accent sur le recueil d'informations comme stratégie préalable à l'activité ou comme moyen de sensibilisation, voire d'apprentissage.

L'action préventive sanitaire a été introduite en cours d'analyse pour tenir compte des démarches individualisées adressées à des personnes à risques dans le cadre de programmes très globaux, par exemple des consultations diététiques.

Les catégories de ressources n'appellent pas de commentaires particuliers si ce n'est la catégorie "gestion globale" qui fut aussi introduite en cours d'analyse. Elle ne concerne pas la gestion quotidienne du programme, mais plutôt l'organisation générale des relations entre les différents acteurs du programme.

V. LA THESE

L'évaluation est une activité de nature unique qui comporte trois composantes, toujours présentes même si elles sont parfois cachées ou différées : un processus, un jugement et une décision.

L'unicité profonde de l'évaluation permet donc de rassembler toutes les pratiques d'évaluation dans un modèle commun.

L'énonciation sociale, stratégique et technique de l'une ou de plusieurs de ces trois composantes détermine les différentes pratiques d'évaluation que l'on peut rencontrer aussi bien dans la littérature que sur le terrain.

Cette différenciation des fonctions permet de comparer et de comprendre les modèles théoriques développés au fil du temps. Elle permet de rendre compte des pratiques d'évaluation de programmes rencontrées en éducation pour la santé en Communauté française de Belgique.

La combinaison de ces fonctions et de ces composantes offre 124 modalités théoriques d'évaluation.

Nous posons l'hypothèse que les pratiques d'évaluation de programmes se caractérisent par une utilisation différente des modalités d'évaluation en fonction de la structure et du contenu du programme.

CHAPITRE 2

LES MODELES ET PARADIGMES D'EVALUATION

I.	L'EVALUATION "DECISION"	50
II.	L'EVALUATION D'EXPERTISE	53
II.A.	LA CLASSIFICATION DE HOUSE	54
II.B.	LE PARADIGME D'EXPERTISE	56
II.C.	LE PARADIGME DE PRETOIRE	57
II.D.	LES PRATIQUES DE META-EVALUATION	58
III.	L'EVALUATION FORMELLE	62
III.A.	L'EVALUATION COMME PRODUCTION DE CONNAISSANCE.....	62
III.B.	L'EVOLUTION DES PARADIGMES DE L'EVALUATION SCIENTIFIQUE	63
III.C.	LES MODALITES DE L'EVALUATION FORMELLE	72
IV.	L'EVALUATION DE CONTROLE	74
IV.A.	LE PARADIGME JURIDIQUE	74
IV.B.	LE PARADIGME TECHNICO-ECONOMIQUE	75
IV.C.	LES LIENS AVEC L'EVALUATION D'EXPERTISE	75
IV.D.	LES MODALITES DE L'EVALUATION DE CONTROLE	76
V.	L'EVALUATION ORIENTEE VERS LES CONCLUSIONS	78
V.A.	SCRIVEN.....	78
V.B.	STAKE.....	81
V.C.	CRONBACH.....	84
V.D.	LES IDEOLOGIES DE L'EVALUATION	86
V.E.	LES MODALITES DE L'EVALUATION ORIENTEE VERS LES CONCLUSIONS	88
VI.	L'EVALUATION "PROCESSUS - DECISION"	90
VI.A.	LES DISTINCTIONS ENTRE L'EVALUATION FORMATIVE ET SOMMATIVE.....	91
VI.B.	LES DIFFERENTS MODELES THEORIQUES ET L'EVALUATION ORIENTEE VERS LA DECISION	92
VI.C.	L'APPROCHE "DECISIONNELLE" DE STUFFLEBEAM	94
VI.D.	L'APPROCHE "CENTREE SUR L'UTILISATION"	95
VI.E.	L'APPROCHE "PAR LES ENJEUX"	97
VI.F.	LES MODALITES DE L'EVALUATION "PROCESSUS - DECISION"	101
VII.	L'EVALUATION TOTALEMENT EXPLICITE	103
VII.A.	EVOLUTION DE LA DEFINITION DE L'EVALUATION DE STUFFLEBEAM	103
VII.B.	MONNIER.....	103
VII.C.	VIAL.....	103
VII.D.	LES MODALITES DE L'EVALUATION EXPLICITE	105

La grille APEP postule l'existence d'une unité de nature entre les pratiques d'évaluation de programmes. Les différences entre les pratiques seraient fonction du contexte d'utilisation de l'évaluation.

L'examen des modèles d'évaluation proposés par de nombreux auteurs et des paradigmes utilisés depuis le siècle dernier offre une occasion de valider la grille d'analyse.

Un "modèle" est un ensemble cohérent de propositions théoriques et pratiques construit par un auteur à un moment donné et "donné en exemple". Il s'agit le plus souvent de la théorisation d'une pratique et non des résultats d'une recherche. Les recherches sur l'évaluation seront examinées dans le chapitre suivant. Par "paradigme" ou "approche", nous entendons un ensemble théorique et pratique plus large et souvent plus durable qui regroupe des "modèles" ou les oriente. Ces modèles et ces approches permettent d'affiner la compréhension des réponses apportées par les évaluateurs aux multiples situations qu'ils rencontrent.

La présentation de ces modèles et paradigmes se fera à partir de quelques ouvrages de synthèse. C'est-à-dire d'un certain nombre d'auteurs qui ont tenté d'analyser, de comparer et de décrire l'évolution historique de l'évaluation de programmes à partir de textes originaux. Cette option se justifie par des raisons essentiellement pratiques. La littérature sur l'évaluation, presque essentiellement américaine, est très abondante et souvent très redondante. Un retour systématique aux sources des principaux auteurs ne manquerait certes pas d'intérêt, mais dépasserait largement le but de ce travail. De plus, notre préoccupation procède à la fois de la recherche d'une grande diversité et de la construction d'une vision d'ensemble. Certains modèles particulièrement intéressants ont néanmoins été approfondis à partir des textes initiaux.

Pour chacun des types d'évaluation prévu par la grille, nous identifierons le ou les modèles qui semblent y correspondre en le justifiant. Nous les situerons également dans une perspective historique. L'évolution d'une pratique professionnelle est une réponse aux changements sociaux, aux évolutions techniques et théoriques. Nous tenterons de le montrer.

Les classifications d'évaluation sont proposées à partir des ouvrages et articles de Alkin et Ellett (1990), House (1978, 1986), Monnier (1987), Scheerens (1990), Stufflebeam et Webster (1980), Stufflebeam et Shinkfield (1985), Worthen (1990).

La mise en perspective historique des pratiques d'évaluation s'appuie sur les textes de Barbier (1990, pp. 36 à 61), Desrosières (1993, pp. 203 à 211), Madaus, Stufflebeam et Scriven (1983, pp. 3 à 23), Monnier (1987, pp. 17 à 36), Stufflebeam et Shinkfield (1985).

I. L'ÉVALUATION "DECISION"

T1 : L'évaluation "décision" (décision explicite)

		Décision	<i>Explicite</i>
Processus	Jugement de valeur		<i>Implicite</i>

Dans la vaste littérature que nous avons consultée, Barbier (1990) est le seul à considérer clairement les usages du jugement de valeur, ses effets, comme une composante de l'évaluation. Pourquoi ?

Barbier perçoit l'universalité du rôle de l'évaluation dans la conduite de l'action. *"On peut penser que l'immense majorité des processus sociaux d'évaluation s'effectue de cette façon et que le repérage des autres formes de manifestation de l'évaluation doit se faire sur le fond de celle-ci."* (p. 32) Les décisions d'action sont fonction de jugements de valeur sur la conduite à tenir.

Si Barbier prend comme point de départ la décision, c'est parce qu'il s'appuie sur l'analyse des évaluations d'actions sociales : relations écrites ou orales de pratiques d'évaluation de formation.

Cette démarche n'est pas celle des auteurs américains que nous avons lus. Ceux-ci partent de leurs propres expériences de professionnels de l'évaluation. Cette position les conduirait d'emblée à considérer l'évaluation, non pas comme un phénomène continu, mais discontinu. L'évaluateur professionnel est extérieur à l'action décidée et conduite par d'autres acteurs. La position des différents acteurs, leur "spécialisation" comme évaluateur, juge ou décideur vis-à-vis du programme permet de comprendre l'apparente discontinuité de l'activité évaluative. Cette séparation des fonctions sera d'ailleurs d'autant plus grande que les programmes seront de plus grande ampleur et mobiliseront plus de personnes. Plus qu'un modèle théorique aussi précis soit-il, c'est l'analyse et l'observation qui permettront le repérage des fonctions de chacun. Nous considérons à chaque fois les pratiques d'évaluation en regard de l'unité d'action que constitue le programme plus que par rapport à une unité de temps ou de lieu.

Mais à l'intérieur de cette unité "programme", de multiples cycles d'action peuvent coexister. Par cycle, nous entendons la conduite d'une action selon le schéma présenté au chapitre 1 (IV.A.) comprenant différentes étapes articulées les unes aux autres par des processus d'anticipation et de rétroaction. Les cycles courts et complets menés par une seule personne auront des modalités d'évaluation différentes des cycles plus longs ou incomplets.

Les 4 modalités de l'évaluation "décision"

	INTERNE			EXTERNE		
INFORMEL	P	J	Dii	P	J	Dei
FORMEL	P	J	Dif	P	J	Def

Sur cette base, ce sont les modalités de l'évaluation "décision" interne, formelle et informelle, qui seraient les plus nombreuses. Les décisions formelles (P J Dif) concerneraient des actes de gestion quotidienne, directs, immédiats. Mais l'énonciation de ces pratiques, parce que quotidiennes, pourrait ne pas être aussi fréquente. A priori, le "besoin" social d'une telle énonciation est faible. Les décisions informelles (P J Dii) seraient des décisions "stratégiques" aux effets plus lointains et aux sources plus diverses.

Les modalités externes (P J Def et P J Dei) seraient des décisions externes au programme qui l'influenceraient directement ou indirectement, comme les décisions de nature politique ou administrative. A cet égard, le recours de plus en plus fréquent à une gestion sociale par programmes rapproche de fait la décision politique de la gestion plus directe, entraîne des rapports différents entre acteurs et le recours plus fréquent à d'autres types d'évaluation plus développés. L'évaluation des politiques (Nioche, 1982) signifiera dès lors l'évaluation des programmes d'action décidés et menés par les pouvoirs publics et ne devra pas être confondue avec l'évaluation de la politique considérée comme domaine d'activité spécifique.

Le rapport de l'évaluation au décideur

Le regard de théoriciens de l'évaluation sur leur rapport au politique ou plus largement au décideur mérite un petit détour.

La classification de Stufflebeam et Webster (1980) porte sur treize approches de l'évaluation identifiées à partir des travaux de Stake, Hastings et Guba.

Pour ces auteurs, l'évaluation est une activité construite et conduite pour assister un public donné dans le jugement et l'amélioration de la valeur d'un objet éducatif donné. Le but assigné à l'évaluation doit être d'apporter une appréciation sur la valeur d'un objet ou d'une action. Sur cette base, les auteurs créent trois catégories : les fausses évaluations, les quasi-évaluations et les vraies évaluations.

Les fausses évaluations, qui nous intéressent ici, ne tiennent pas compte de la valeur réelle d'une activité et ne sont que des prétextes à d'autres buts. Ce sont les études contrôlées politiquement qui ont pour objet de maintenir ou d'augmenter l'influence du commanditaire de l'évaluation. Ce sont aussi les études inspirées par les relations publiques pour créer une image positive d'un programme.

Dans une présentation des démarches pré-évaluatives, Patton (1978) mentionne aussi la démarche à but politique, qui est, selon lui, très fréquente dans le secteur social. L'évaluation "politique" ne s'occupe pas de l'efficacité réelle du programme, mais de son efficacité politique. Seule la décision compte.

Weiss et Bucuvalas (cité par Alkin, 1990) concluent que les évaluations dont la méthodologie est remise en cause par les décideurs sont celles dont les résultats ne confortent pas une position préétablie.

Alkin et Coyle (1988) répertorient différents types de mauvais usages de l'évaluation et citent l'absence volontaire d'utilisation et le détournement des conclusions.

Peut-on rejeter ces modes d'évaluation et les qualifier de faux, de préévaluatifs ou de mauvais usages ? Ils ne s'accordent pas avec le modèle d'évaluation de ces auteurs, à la recherche de la valeur "objective" d'une action, mais ils correspondent à des pratiques courantes. Leur simple existence justifie de les considérer comme des pratiques évaluatives à comprendre. Il s'agit d'activités construites et conduites pour assister un public donné à juger de la valeur d'un objet éducatif donné. Cronbach (1980) insiste sur la composante sociale et politique de l'évaluation et l'absence d'une vérité indépendante des enjeux et des contextes.

La persistance des pratiques évaluatives de ce type suggère que l'évaluation poursuit d'autres buts que la recherche de la "vérité" d'un programme. *"Malgré les déceptions et les échecs, la demande (d'évaluation) reste très forte montrant que les résultats explicites recueillis ne sont qu'une partie des effets qu'on en attend."* (Barbier, 1990, p. 58) Ces autres effets s'analysent en termes de contrôle et de pouvoir.

Remarquons qu'une fracture apparaît entre la décision et les deux autres composantes de l'évaluation. L'examen de l'évaluation "processus - décision" (type VI) fournira l'occasion de citer bien d'autres exemples sur le rapport entre le processus d'évaluation, le jugement de valeur et la décision.

II. L'EVALUATION D'EXPERTISE

T2 : L'évaluation d'expertise (jugement explicite)

Jugement de valeur		<i>Explicite</i> <i>Implicite</i>
Processus	Décision	

Les 4 modalités de l'évaluation d'expertise

	INTERNE			EXTERNE		
INFORMEL	P	Jii	D	P	Jei	D
FORMEL	P	Jif	D	P	Jef	D

Selon Barbier (1990), l'expression spontanée d'un jugement de valeur sur une activité est une évaluation. Ce type d'évaluation s'effectue très fréquemment par rapport à des critères qui ne sont pas explicités, sans recours à des procédures méthodologiques. Selon cet auteur, ces évaluations ne sont pas socialement organisées, ni formulées par des personnes socialement déléguées pour le faire. Elles jouent cependant un rôle soit pour influencer ou déterminer les décisions, soit pour servir d'informations de base dans des évaluations méthodologiquement construites. Cette description correspond à ce que nous appelons une évaluation d'expertise menée par un acteur interne à l'action sans formalisation des critères de jugement (P Jii D).

Ce jugement de valeur "spontané" est mal considéré par les professionnels de l'évaluation et souvent aussi par les responsables de décisions. Le jugement de l'expert de terrain n'aurait-il pas de valeur ? Selon ces professionnels, il n'aurait en tout cas pas de valeur objective, sociale.

Une première forme de "valorisation" sociale du jugement interne est la formalisation des critères, soit P Jif D, la deuxième modalité prévue dans la grille d'analyse pour ce type II.

La formulation opérationnelle des objectifs d'un programme par ses acteurs est une manière de formaliser les jugements de valeur internes. Ces acteurs peuvent soit construire leurs propres critères, soit choisir des critères extérieurs standardisés ou normes. C'est le fait de choisir ces normes qui maintient le jugement interne. Cette opérationnalisation sera probablement rarement spontanée, même si elle reste interne. Les raisons seront à rechercher dans les stratégies des acteurs et de l'organisation. Ces stratégies peuvent bien entendu varier au sein d'une même organisation.

Le jugement de valeur externe semble aussi fréquent et particulièrement organisé sur le plan social, contrairement à ce qu'affirme Barbier. C'est pourquoi nous avons qualifié "expertise" les modalités regroupées dans ce type d'évaluation. L'expertise est en effet le prototype de l'évaluation basée sur le jugement de valeur. C'est une des approches évaluatives historiquement les plus anciennes. On la retrouve également dans la plupart des taxonomies de l'évaluation.

Nous développerons ici les modèles d'évaluation externe axés principalement sur le jugement de valeur (l'expertise, l'accréditation, le consensus). A la charnière entre le processus d'évaluation et la décision, le jugement de valeur sera souvent associé à l'un ou à l'autre dans son énonciation. Ces associations seront consultées plus loin respectivement dans les types IV et V.

- Le décideur politique ou administratif, interne ou externe, explicite son jugement et sa décision. Nous développerons cet aspect dans le quatrième type d'évaluation (l'évaluation de contrôle).
- L'évaluateur professionnel ou le chercheur va, lui, expliciter le jugement de valeur en même temps qu'il énonce le processus d'évaluation. Au delà de son rôle technique, l'évaluateur intervient dans la formulation du jugement quand il ne l'assume pas entièrement. C'est le type VI de notre grille (l'évaluation orientée vers les conclusions).

Plusieurs auteurs (Alkin et Ellett, 1990, Monnier, 1987, Worthen, 1990) intègrent l'expertise professionnelle dans leur classification des différentes formes d'évaluation.

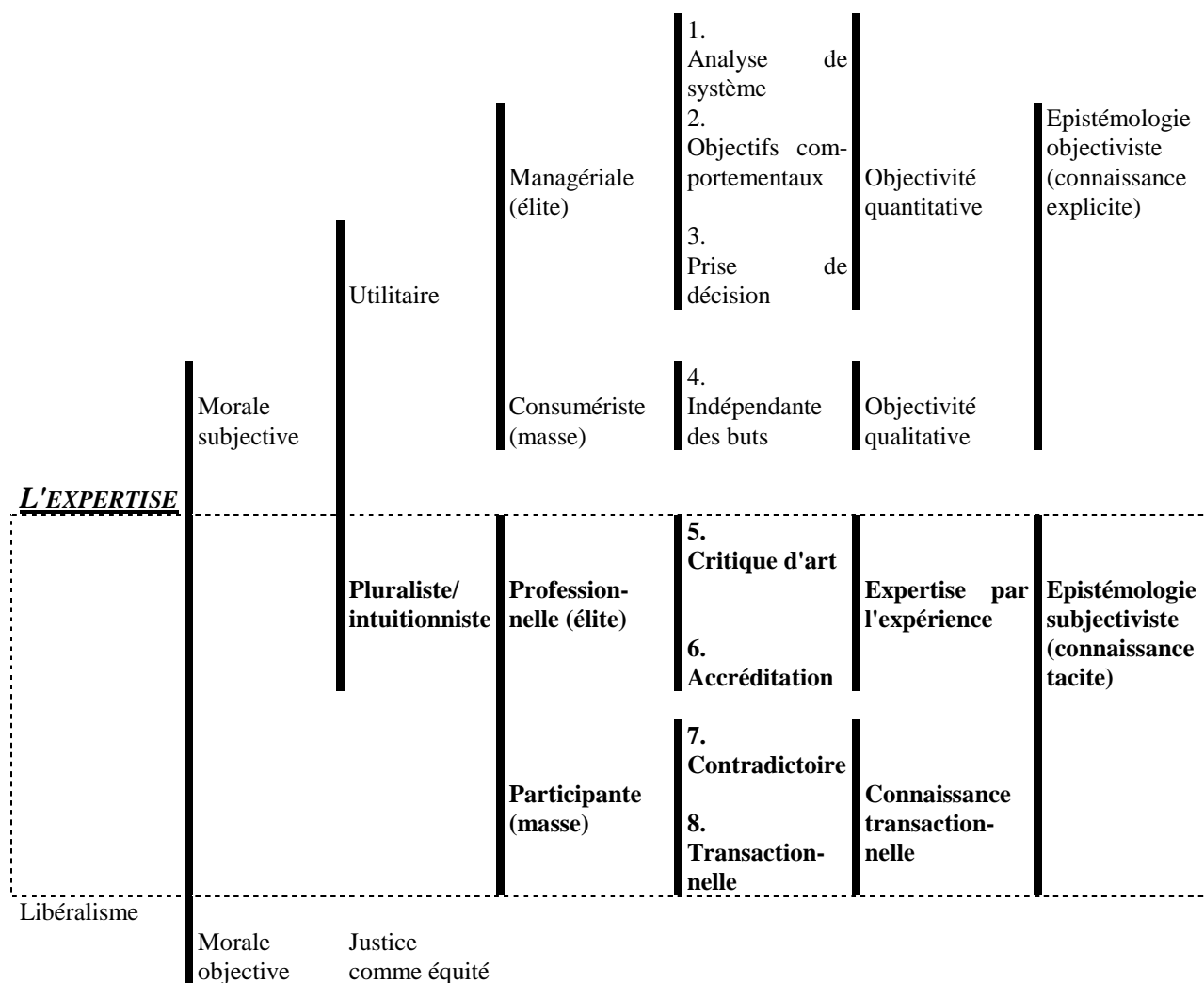
II.A. LA CLASSIFICATION DE HOUSE

Les modèles de House (1978) offrent des pistes pour distinguer différentes formes d'évaluation d'expertise.

House analyse et compare huit modèles d'évaluation en identifiant les hypothèses théoriques sous-jacentes. Il propose tout d'abord deux critères, le libéralisme et la morale subjective, qui fondent l'évaluation. Il identifie ensuite deux groupes sur la base à la fois d'un critère moral et d'un critère épistémologique. Le premier regroupe des modèles qui visent l'utilité sociale (assurer un maximum de bonheur à la société), qui se réfèrent à une épistémologie objectiviste à la recherche d'une connaissance explicite. Nous n'envisagerons pas ici ces modèles qui explicitent leur processus d'évaluation.

Le second groupe (encadré dans le schéma suivant) comprend des modèles axés sur le pluralisme des critères moraux, en référence à une épistémologie également subjectiviste basée sur la connaissance tacite.

Classification de House (1978)



La critique d'art (5) se base sur la formation et l'expérience d'un expert externe à l'action.

L'accréditation (6) s'appuie aussi sur l'expérience de professionnels externes utilisant des normes d'analyse.

L'approche contradictoire (7) utilise des procédures de type judiciaire pour examiner contradictoirement les aspects négatifs et positifs d'un programme, le plus souvent sous la forme d'un jury.

L'approche transactionnelle étudie l'action par des méthodes informelles comme l'étude de cas.

Ces quatre modèles forment deux groupes distincts par le public visé : une élite professionnelle (décideurs, gestionnaires) ou un public plus large d'utilisateurs et de praticiens. D'un côté l'expertise est basée sur l'expérience subjective, de l'autre sur la confrontation des subjectivités (connaissance transactionnelle).

Pour House, même s'il existe d'autres modèles d'évaluation, ils peuvent être assimilés à ces quatre exemples s'ils répondent aux mêmes critères.

Quel lien voyons nous avec nos propres critères ?

La distinction de House entre une épistémologie "objective" et "subjective" pourrait tracer la ligne marquée par l'explicitation du processus, bien que nous ne partagions pas entièrement cette analyse. En effet, Bryman (1984) indique la rupture possible entre épistémologie et méthodologie.

Les quatre modèles se situent dans une démarche d'expertise externe. Selon nous, ce qui peut les différencier, c'est autant la formalisation du jugement de valeur que le public concerné. Examinons plus avant ces quatre modèles.

En s'inspirant de la classification de Guba et Lincoln, Monnier (1987) propose une typologie des principaux paradigmes qui sous-tendent les modèles d'évaluation. Il cite notamment le paradigme d'expertise et le paradigme de prétoire.

II.B. LE PARADIGME D'EXPERTISE

Selon Monnier, dans le paradigme d'expertise, la vérité est fondée sur la conviction d'une personne reconnue pour son savoir et ses compétences. Mais c'est la connaissance exogène et non subjective qui construit la pensée de l'expert. Ce type d'évaluation est fréquent dans les audits, dans l'évaluation des activités scientifiques (au sein de la communauté des chercheurs) ou dans les démarches d'accréditation.

C'est aussi un paradigme ancien. L'accréditation, par exemple, prend ses racines aux Etats-Unis à la fin du 18^{ème} siècle (Madaus, Stufflebeam et Scriven, 1983). Elle reste toujours une méthode très employée pour l'évaluation des institutions sociales, scolaires ou sanitaires. L'évaluation d'expert est aussi une forme reconnue d'évaluation de programmes dès le début du 20^{ème} siècle dans le même pays ; souvent en complément des méthodes psychométriques de mesures du rendement scolaire.

Il semble possible de distinguer ces différentes formes d'expertise sur l'axe de la formalisation :

P Jei D : une expertise intuitive, sans critères formels ou une expertise avec des critères formalisés mais implicites (ils peuvent être découverts ou clarifiés a posteriori) ;

P Jef D : une expertise s'appuyant clairement sur des normes reconnues ou du moins communiquées.

On voit que la tâche est délicate. Certains modèles s'inscrivent plus nettement que d'autres dans ce schéma.

Eisner (1975) suggère un modèle d'expertise, de jugement professionnel intuitif assez radical. L'expert se base sur ses connaissances, son expérience personnelle. Les facultés d'appréciation de l'œnologue et l'analyse du critique d'art sont proposées comme modèles d'une connaissance éducative et d'une critique éducative. L'évaluateur doit découvrir les subtiles nuances de l'acte éducatif comme l'œnologue et être capable d'exprimer celles-ci comme le critique d'art. L'appréciation est intime, privée et la critique est externe, publique. A Stanford, le travail d'Eisner consiste à clarifier les concepts, les critères, les techniques de cette capacité d'apprécier et de restituer l'acte éducatif. Bien que radical, ce modèle est souvent celui des experts. Le modèle d'expertise de Eisner tout comme l'évaluation d'accréditation sont qualifiés d'évaluation "vraie" par Stufflebeam et Webster (1980).

L'accréditation nous paraît l'exemple de l'expertise professionnelle externe formalisée distincte de l'expertise intuitive de Eisner. Les normes y sont construites, spécifiées. Les experts jugent en fonction de ces normes.

II.C. LE PARADIGME DE PRETOIRE

Dans le paradigme de prétoire (Monnier, 1987), la vérité émerge de la confrontation des points de vue, de l'exposé de thèses opposées, de la confrontation de preuves contradictoires face à des instances dénuées de préjugés. C'est le modèle de la procédure judiciaire ou des enquêtes parlementaires. Nous sommes dans un type d'évaluation où l'expertise externe est plurielle.

Madaus, Stufflebeam et Scriven (1983) signalent l'apparition dès le 19^{ème} siècle de commissions parlementaires chargées d'examiner l'efficacité et de juger la qualité des systèmes sociaux mis en place. Ces commissions recueillent des témoignages, et les confrontent pour proposer des solutions. Ils nous proposent deux exemples. La *Royal Commission of Inquiry into Primary Education* en Irlande conclut en 1870 que les progrès des enfants ne sont pas ce qu'ils devraient être et propose de payer les enseignants en fonction des résultats de leurs élèves en lecture, orthographe, écriture et arithmétique. La *Royal Commission on Small Pox and Fever Hospitals* en Angleterre recommande en 1882 l'ouverture des hôpitaux à tous les citoyens.

De telles commissions existent toujours et restent une des voies de l'évaluation des politiques. Ces auteurs les qualifient d'évaluation symbolique et emblématique, de pseudo-évaluation.

Dans les années septante, ce paradigme ancien a été repris aux Etats-Unis comme une technique d'évaluation des programmes d'éducation : l'approche contradictoire (*adversary*). Deux équipes d'évaluateurs sont chargées d'explorer les arguments en faveur et en défaveur d'un programme pour fournir à des décideurs la meilleure information possible pour juger ou pour sélectionner. Cette démarche suit un processus dialectique et légal. Elle a été proposée par Owens en 1973 et formalisée particulièrement par Wolf la même année dans son modèle "judiciaire". L'évaluateur y devient un avocat chargé de présenter ses "preuves" à un jury devant un juge. On sait que la procédure judiciaire sépare la constitution du dossier à charge et à décharge, le jugement et l'application des peines, soit nos trois composantes. Mais, dans le cadre de la justice, le jugement est établi par rapport à des normes, des lois.

Grâce aux moyens scientifiques de détermination des preuves lors des enquêtes, certains auteurs comme Wolf (Alkin et Ellett, 1979) rejettent le qualitatif d'évaluation de prétoire. Wolf l'assimile plutôt à un paradigme scientifique où la procédure n'est qu'une mise en scène. La vérité scientifique basée sur des faits doit sortir victorieuse de la confrontation. Il désavoue l'argumentation et la persuasion. Pour d'autres, comme Rippey (Alkin et Ellett, 1979, Worthen, 1990), l'argumentation lors du jugement reste au coeur de la démarche. Il développe un modèle d'évaluation "transactionnelle" où l'évaluation devient une stratégie de gestion des conflits.

Les conférences de consensus, fréquentes dans le domaine de la santé publique, peuvent être assimilées à ce type d'évaluation et ouvrent sur le même débat. Nous en parlerons dans le chapitre 4.

Worthen (1990) fait quant à lui la distinction entre une approche "contradictoire" et "pluraliste - intuitionniste". La première organise la confrontation de points de vues divergents par différents types de procédures (jury, consensus) pour arriver à un seul avis, le meilleur. La seconde admet l'existence de différentes valeurs et besoins et tente de rejoindre cette pluralité dans une forme globale et intuitive.

L'évaluation d'expertise montre la diversité des formes d'évaluation et la difficulté de regrouper ou de différencier les modèles. Ce sont plus souvent les conditions d'application de ces modèles en situation réelle qui permettent de situer les pratiques. La position interne ou externe du ou des acteurs, la formalisation ou non des critères de jugement permet, dans ces limites, de clarifier certains modèles théoriques.

II.D. LES PRATIQUES DE META-EVALUATION

Les différentes pratiques de méta-évaluation apportent un éclairage intéressant sur l'évaluation d'expertise, en fournissant une transition vers l'évaluation formelle (type III).

Scriven crée le terme "méta-évaluation" en 1967 (Stufflebeam et Shinkfield, 1985). Pour Scriven, l'évaluation est avant tout la production d'un jugement de valeur. C'est l'évaluateur qui est responsable de ce jugement face à la société. Il va dès lors travailler à la professionnalisation de cette activité en créant des revues, des associations, en développant des "normes" professionnelles et la notion de méta-évaluation. Celle-ci pourra soutenir la conduite d'une évaluation de qualité, mais aussi permettre de juger de la compétence technique et de la validité des conclusions. C'est une évaluation des pratiques d'évaluation.

Le but est clairement d'émettre un jugement de valeur sur le processus d'évaluation et sur son produit. A partir de ce but initial, on verra se développer deux courants, celui des normes professionnelles, qui touche surtout à l'évaluation de type expertise, et celui de la méta-analyse (Straw et Cook, 1990) centré sur les revues de littérature et l'analyse statistique d'un ensemble de données, qui touche donc davantage à l'évaluation de type formel.

Les normes professionnelles ou "Standards" constituent une étape importante dans la professionnalisation de l'évaluation.

Ce sont les "Standards" définis par un comité présidé par Stufflebeam qui servent le plus souvent de référence (*Joint Committee on Standards for Educational Evaluation*, 1981). Une trentaine de critères ont été établis et regroupés en quatre catégories :

- utilité : répondre aux besoins pratiques d'information ;
- faisabilité : réaliser une évaluation réaliste et prudente ;
- honnêteté : suivre une ligne légale et éthique ;
- précision : recueillir des informations techniquement correctes.

Le but des "Standards" est d'améliorer l'évaluation et d'informer sur son efficacité.

Les "Standards" ont été mis au point à partir d'un processus qui dura près de cinq ans : rassemblement de la littérature, liste de cas, d'erreurs ; puis rédaction des "Standards" par une trentaine d'experts indépendants, sélection dans un premier rapport ; examen et critique par 50 personnes désignées par des organisations, publication finale.

Le format utilisé pour la présentation des "Standards" montre le sérieux de la tentative : descripteurs, définitions, lignes de conduite pour la procédure à suivre, erreurs communes, précautions, illustrations comprenant la violation du critère et des mesures correctives possibles.

A ce stade, nous retiendrons essentiellement la dimension éthique comme caractéristique du jugement d'expertise. Les autres dimensions font référence aux pôles technique (faisabilité et précision) ou décisionnel (utilité).

Les Standards de rectitude

Ils sont prévus pour garantir qu'une évaluation sera conduite légalement, avec éthique, en respectant le bien-être de ceux qui sont impliqués dans l'évaluation, aussi bien que le bien-être de ceux qui sont affectés par ses résultats.

1. engagement formel
2. conflit d'intérêt
3. divulgation pleine et entière
4. droit du public de savoir
5. droits de l'homme
6. relations humaines
7. équilibre du rapport
8. responsabilité financière

Les "Standards" ne veulent pas privilégier une approche de l'évaluation plutôt qu'une autre, mais veulent encourager la variété et la diversité : survey, revue de documents, jury, plaidoirie, test, étude de simulation, check-list, schéma quasi-expérimental, méta-analyse, analyse de cas, etc. Ils tiennent compte des apports de différents courants théoriques comme de la dimension politique.

Les "Standards" ne doivent pas être vus comme un exercice académique pour les évaluateurs professionnels, mais servir de code pour améliorer la pratique de l'évaluation à tous les niveaux : des plus petites aux plus grandes, des plus formelles aux moins formelles. Ce message est applicable à tous les évaluateurs et à ceux qui utilisent les services d'évaluateurs. En fait, tous les critères ne doivent pas être rencontrés de manière systématique, mais servir de points de repère.

De même certains critères peuvent entrer en conflit, par exemple la validité et la fidélité des informations avec les limites de coût. L'évaluateur devra assurer un équilibre judicieux en fonction de l'évaluation à mener. Le Comité offre cependant une direction à suivre en présentant les quatre catégories dans un ordre déterminé : l'utilité, la faisabilité, l'honnêteté et la validité. La préoccupation pour une information techniquement correcte passe après l'assurance de l'utilité, de la faisabilité ou du caractère éthique.

Le Comité fournit également une table permettant d'identifier les critères particuliers les plus importants à différents stades d'une évaluation. L'évaluateur peut ainsi s'interroger sur la pertinence de certains choix : décider de faire l'étude ; clarifier les objectifs ; assurer sa viabilité ; signer le contrat ; constituer l'équipe ; gérer l'étude ; collecter les données ; les analyser, les rapporter ou appliquer les résultats. Un formulaire d'engagement écrit sur les "Standards" est aussi proposé.

Les "critères professionnels d'une évaluation éducative" ont suscité un grand intérêt aux Etats-Unis et dans d'autres pays. Ils ont été jugés utiles sur le plan pratique, car ils étaient applicables à de nombreuses situations d'évaluation. Ils ont servi pour construire des projets d'évaluation, juger des propositions d'évaluation, juger des rapports, assurer des formations.

C'est un bel exemple d'évaluation d'expertise externe formalisée. La procédure a pour but de construire les normes. C'est le jugement de valeur qui est l'axe essentiel de cette méta-évaluation. Les "Standards" sont un support pour des experts externes, mais ils peuvent aussi être utiles pour des évaluation internes. Cependant Worthen (1990) remarque que les méta-évaluations externes utilisant les "Standards" restent rares.

III. L'ÉVALUATION FORMELLE

T3 : L'évaluation formelle (processus explicite)

Processus			<i>Explicite</i>
	Jugement de valeur	Décision	<i>Implicite</i>

En 1988, Alkin (1990), directeur du Centre pour l'étude de l'évaluation à l'université de Californie, réunit plusieurs de ses collègues (Patton, Weiss, House, McLaughlin, Conner, Kean, King, Law) pendant trois jours pour débattre de manière informelle de l'évaluation et des problèmes qu'elle soulève.

Au centre de ce débat, on trouve la ou les fonction(s) de l'évaluation. L'objectif fixé à l'évaluation oriente largement les conceptions que se fait l'évaluateur de son rôle, du type d'évaluation qu'il va mener et des relations qu'il va établir avec ses différents publics.

Pour ces spécialistes, le but principal de l'évaluation est de produire de la connaissance sur les pratiques. C'est aux utilisateurs de l'évaluation qu'incombe la responsabilité d'utiliser cette connaissance pour devenir plus efficaces, pour améliorer leur pratique.

III.A. L'ÉVALUATION COMME PRODUCTION DE CONNAISSANCE

Cette production de connaissance peut être vue, selon eux, de deux points de vue différents.

D'une part, il y a une connaissance absolue, irréfutable, capable d'être généralisée. Elle est produite par un travail systématique et scientifique. Il ne peut y avoir alors de distinction entre la recherche et l'évaluation. C'est la position de Weiss, McLaughlin et Conner.

Le chercheur se sent responsable d'explorer tous les aspects d'un programme, notamment en s'appuyant sur la littérature, il ne se limite pas aux sujets souhaités par l'utilisateur. Les connaissances théoriques doivent être utilisées pour explorer les variables critiques. L'évaluateur choisit ces variables en fonction de la situation et de sa propre expérience.

Le public de l'évaluateur est nécessairement large, au-delà du commanditaire ; il comprend la communauté scientifique et plus largement la population. L'évaluateur a une responsabilité de diffusion de la connaissance de manière large comme s'il s'agissait d'une recherche.

D'autre part, il y a une connaissance relative. Elle est orientée vers la résolution de problèmes particuliers, prioritairement au service des utilisateurs directs comme les praticiens, les directeurs de programme, les décideurs politiques. C'est la position de Patton, Alkin et King.

Selon Patton, l'évaluation doit répondre "aux utilisateurs désignés pour des usages définis" ("*intended user for intended use*").

L'évaluateur aide l'utilisateur à découvrir ses propres questions. Il peut aussi guider le client vers les questions les plus pertinentes en relation avec les connaissances existantes. La connaissance pratique du client est une théorie qui est analysée et testée parmi d'autres, en accord avec celui-ci.

La responsabilité de l'évaluateur se limite à son public direct, elle ne s'étend pas aux usages secondaires de l'évaluation. Ceux-ci ne sont pas un objectif prioritaire et ne sont pris en compte de façon active qu'à la demande des clients. Des considérations pratiques de ressources limitent de toute façon le champ des possibilités.

Le bref compte rendu de ce débat nous indique que les divergences entre ces professionnels concernent la valeur des conclusions et leur usage. Nous resituerons ces options dans l'examen des types d'évaluation orientés vers les conclusions ou les décisions (types V et VI). Mais le point de départ est unanime : produire des connaissances.

Ce n'est pas la position défendue par tous les évaluateurs. D'autres situent l'enjeu principal dans le jugement de valeur. Ils établissent une séparation nette entre la recherche qui produit des connaissances et l'évaluation qui recueille des informations tout en établissant des critères pour produire un jugement de valeur (Barbier, 1990).

Recherche, recherche évaluative, évaluation, nombreux sont les auteurs qui tentent d'établir des lignes de démarcation (Barbier, 1990, Suchman, 1967, Worthen, 1990). Au delà des termes et des lieux de rupture, ce qui importe ici, c'est de reconnaître l'ancrage de l'évaluation dans le développement explicite et formel de son processus.

Avant d'examiner les diverses modalités de l'évaluation explicite formelle, il est utile d'en comprendre l'évolution historique et les différents paradigmes.

III.B. L'EVOLUTION DES PARADIGMES DE L'EVALUATION SCIENTIFIQUE

1. Le premier paradigme de l'évaluation "scientifique" est probablement statistique.

Dans son histoire de la raison statistique, Desrosières (1993) offre quelques points de repère sur les relations entre l'examen macroscopique et quantitatif des phénomènes sociaux et leur gestion politique.

Le 18^{ème} siècle a vu l'apparition des états. Le 19^{ème} se caractérise par l'émergence de systèmes sociaux et sanitaires de plus en plus structurés. En Angleterre, le *General Register Office* (GRO), système d'enregistrement des statistiques sociales, naît en 1837. Dirigé jusqu'en 1880 par le médecin William Farr, il va jouer un rôle important dans le développement de la santé publique en centrant l'attention et les débats sur l'environnement économique et social comme facteur explicatif de la mortalité. Les taux de mortalité sont inscrits dans une loi sur la santé publique en 1848. Celle-ci stipule que des réformes sanitaires doivent être entreprises dans les localités dont le taux de mortalité est supérieur à la moyenne nationale calculée par le GRO.

En France, la revue des Annales d'hygiène publique et de médecine légale est créée en 1829 et sera le support d'enquêtes sur les groupes sociaux les plus exposés à la misère, aux maladies. Ces études conduisent à moraliser ces groupes par l'information (l'éducation sanitaire) et à améliorer les conditions de vie par la législation.

Ces recherches constituent les prémises de démarches scientifiques, mais ne sont pas encore de véritables "expérimentations sociales" comme on les rencontrera au 20^{ème} siècle.

Dans le domaine de l'éducation, on peut assimiler à ce paradigme l'accumulation d'informations sur le rendement des étudiants grâce à des tests standardisés. Le baccalauréat naît en 1808 en France. A Boston en 1845, Howe et Mann mettent en place des examens écrits comme en Europe pour remplacer les examens oraux. Une longue tradition américaine d'évaluation des performances des élèves par les tests se met en place. L'évaluation de programmes est déjà bien présente même si elle reste implicite et cachée. En effet, ces deux chercheurs utilisent ces résultats pour comparer les écoles et les maîtres. Ces données servent à modifier les salaires ou à démontrer l'intérêt des punitions corporelles. L'utilisation politique de l'évaluation était déjà bien présente.

Même si l'évaluation n'est pas encore reconnue comme telle, les données descriptives sont utilisées pour expliquer certains phénomènes, attribuer une valeur ou prendre des décisions.

2. Le second paradigme de l'évaluation scientifique est le paradigme expérimentaliste.

a) L'émergence d'une organisation scientifique du domaine social.

Le début du 20^{ème} siècle voit émerger l'idée d'une organisation scientifique du domaine social comme était née la rationalisation du travail dans les entreprises avec Taylor. La quantification et la standardisation deviennent les maîtres-mots. Mais l'évaluation des actions passe encore par l'évaluation des individus.

Rice mène la première évaluation de programme entre 1897 et 1898 (Madaus, Stufflebeam et Scriven, 1983). Il compare la valeur de l'entraînement à l'orthographe à travers les districts scolaires grâce à des scores et des tests. Il plaide déjà pour faire des éducateurs des expérimentalistes quantitatifs et des utilisateurs d'études comparatives sur le rendement des étudiants.

En France, Binet développe en 1904 son test d'intelligence à la suite d'une commande du ministre de l'Instruction publique soucieux de détecter les enfants anormaux afin d'organiser un enseignement qui leur soit adapté (Monnier, 1987). Cet exemple indique l'émergence du paradigme expérimentaliste dans la gestion des affaires publiques. Ce test s'inspire d'une démarche scientifique qui veut appréhender les phénomènes sociaux de manière neutre et objective à l'instar des sciences naturelles.

Les premiers éléments de ce paradigme font leur apparition de manière marquante. Les chercheurs dans le secteur social et éducatif partagent des croyances et des espoirs qui orientent leurs travaux. La mesure des différences individuelles permet d'observer et d'expliquer les phénomènes sociaux. Pour réaliser ces mesures, il est légitime d'utiliser des normes standardisées construites a priori et ayant valeur universelle. La vérité est conçue comme une propriété intrinsèque et invariante de l'univers régi par des lois générales. Ces lois génèrent des hypothèses qui peuvent être vérifiées expérimentalement (Monnier, 1987).

La qualité de l'enseignement, des écoles et des enseignants est ainsi mesurée aux Etats-Unis par des tests "objectifs" de connaissances des élèves en arithmétique, orthographe (Madaus, Stufflebeam et Scriven, 1983). Ces tests sont développés au niveau local dans les districts les plus importants comme à Boston. Des chercheurs comme Thorndike développent des outils sur des objectifs d'instruction précis. Après la première Guerre Mondiale, certains tests deviennent aussi des normes de référence pour comparer les étudiants et les enseignants.

b) L'impulsion de Tyler dans le domaine de l'évaluation.

L'évaluation dans le domaine de l'éducation prend son essor dans les années trente sous l'impulsion notamment de Tyler (Madaus, Stufflebeam et Scriven, 1983). Cet auteur est considéré aux Etats-Unis comme le père fondateur de l'évaluation en éducation.

Tyler publiera en 1942 ses vues sur l'évaluation en éducation et sur la construction de curriculum. Son approche influencera largement la scène scolaire au cours des vingt années qui suivirent.

L'approche de Tyler offre une coupure assez nette avec les pratiques précédentes. L'évaluation est alors centrée uniquement sur les étudiants et sur la mesure de leur succès. Tyler propose de ne plus comparer les résultats des élèves entre eux ou par rapport à une norme, mais de les mettre en relation avec un curriculum et des objectifs comportementaux, avec un programme, avec les moyens investis. L'évaluation de programmes pouvait se mettre en place.

En introduisant les objectifs comme point de référence pour l'évaluation, Tyler quitte en fait l'évaluation de type formel axée uniquement sur l'accumulation des connaissances pour jeter les bases d'une évaluation orientée vers le jugement de valeur. Nous y reviendrons en présentant ce type d'évaluation.

Après la deuxième Guerre Mondiale, aux Etats-Unis, les outils et les techniques de l'évaluation en éducation connaissent un développement considérable, parallèlement au développement explosif du système scolaire : production de tests standardisés, conception de nouveaux plans expérimentaux et de nouvelles procédures statistiques.

Les principes de l'évaluation expérimentale sont unanimement acceptés et ne souffrent pas d'exception. Il faut un traitement inchangé, des échantillons stables composés d'individus tirés au hasard, pour établir des groupes expérimentaux et de contrôle, et des critères de vérification univoques et préétablis.

Les scientifiques accumulent des données, mais la réflexion sur le rôle de l'évaluation est pratiquement inexistante. Les éducateurs ne sont pas encore devenus comptables de leurs efforts. Il n'y a pas de pression externe des états ou du niveau fédéral, peu de moyens financiers sont en jeu. L'évaluation n'est pas mise en relation avec la qualité des services éducatifs et son amélioration.

c) L'explosion de l'évaluation dans les années soixante.

Le lancement de la course vers l'espace et vers la lune va provoquer aux Etats-Unis un investissement important vers les programmes éducatifs, mais aussi une préoccupation vis-à-vis de l'efficacité de cet argent public. L'évaluation devient un passage obligé de tout programme public.

L'évaluation va donc se développer en utilisant les approches développées précédemment, c'est-à-dire :

- l'approche de Tyler pour définir les nouveaux curriculums et surtout formuler les objectifs ;
- des tests nationaux standardisés construits pour refléter les nouveaux objectifs et les nouveaux contenus ;
- l'approche par le jugement professionnel (expertise) pour vérifier périodiquement les efforts des responsables de programmes ;
- des expérimentations sur le terrain pour mesurer l'efficacité de certains programmes.

Bien que le meilleur de la communauté des évaluateurs soit mobilisé avec des moyens financiers importants et toutes les techniques disponibles, l'échec de cette démarche apparaît assez rapidement au niveau local des districts scolaires. Les tests nationaux sont inadaptés pour les situations locales, pour les publics désavantagés et pour le diagnostic et l'amélioration des actions. Ils manquent de sensibilité et de plus, poursuivant des buts commerciaux, ils ne reflètent que les objectifs les plus communs. L'approche expérimentale se révèle aussi peu efficace, car les schémas requis par celle-ci sont inadaptés et trop rigides. Comme l'approche par expertise et expérience ne correspond pas aux exigences de l'administration américaine, les écoles se trouvent sans véritable guide, ni évaluation.

Le résultat s'avère donc décevant et des voix s'élèvent pour mettre en doute l'efficacité de l'évaluation. Cronbach, en 1963, est un des premiers à faire remarquer les difficultés et surtout le manque d'utilité des évaluations construites de manière expérimentale, utilisant des tests normatifs et ne proposant de résultats qu'à la fin des programmes. Cronbach conseille de ne plus évaluer les programmes entre eux, mais de guider leur développement. Il propose de ne plus communiquer aux enseignants les scores globaux ou les moyennes des étudiants, mais leurs résultats à chaque item. Il ouvre ainsi la voie à une série de théoriciens de l'évaluation qui vont mettre en avant l'importance du jugement de valeur et de la décision.

Cependant, le paradigme expérimentaliste poursuit son développement.

Les difficultés de répondre à tous les critères d'une démarche expérimentale conduisent Campbell et Stanley à proposer en 1966 un approfondissement des plans expérimentaux et quasi-expérimentaux : *Experimental and quasi-experimental design for research*, véritable bible de l'expérimentateur social.

Suchman va aussi développer une approche de type expérimental qui influencera les évaluateurs (Stufflebeam et Shinkfield, 1985). Suchman était professeur de sociologie à l'université de Pittsburgh. Il travailla particulièrement dans le champ de la santé publique. Sa contribution à l'évaluation en 1967 (*Evaluative research : principles and practice in public service and social action programs*) en fait un des représentants les plus éminents du mouvement "expérimentaliste" en évaluation. Pour Suchman, l'évaluateur est d'abord et avant tout un chercheur qui utilise la logique de la méthode scientifique.

Mais si cet évaluateur doit travailler avec rigueur, il doit aussi tenir compte de la situation dans laquelle il fonctionne. Le modèle expérimental est l'idéal, mais il faut pouvoir l'adapter. Suchman fait ainsi la distinction entre la recherche de base qui tente de découvrir la connaissance ; la recherche évaluative, qui est une forme de recherche appliquée et qui détermine la mesure avec laquelle un programme atteint les résultats désirés en utilisant les procédures scientifiques de collecte et d'analyse des données ; et l'évaluation qui est le processus qui rend un jugement sur la valeur.

Le paradigme expérimentaliste connaît le maximum de son ampleur à la fin des années soixante avec la mise en place de véritables expérimentations sociales utilisant des ressources humaines, matérielles et financières considérables (Monnier, 1987). L'ambition est de tester "scientifiquement", de mesurer de manière incontestable les effets sociaux de nouvelles formes d'action avant de les généraliser. Une de ces expériences a, par exemple, duré huit ans et coûté 133 millions de dollars. Le paradigme expérimental et la gestion scientifique doivent mener vers une société "expérimentatrice" comme l'a qualifiée Campbell. *"La société "expérimentatrice" est, à mon avis, une alternative possible pour le futur que nous devrions sérieusement envisager. Une représentation de ce futur possible apparaît à l'examen de l'utilisation des sciences sociales par la société et l'Etat. (...) La société "expérimentatrice" serait une société scientifique au sens propre du terme scientifique. Les valeurs scientifiques d'honnêteté, ouverture critique, expérimentation, volonté de faire évoluer les théories établies, seraient autant de règles de conduite."* (cité par Monnier, 1987, p. 50)

Mais ce modèle expérimental "idéal" est progressivement remis en cause : coût excessif, réponses tardives et résultats incertains ou inutilisables. De nouvelles propositions plus attentives à la réalité sociale et aux décisions à prendre vont alors se développer.

d) L'approfondissement des méthodologies de la recherche évaluative.

Aux Etats-Unis, les années septante voient le véritable début de l'évaluation professionnelle et la multiplication des propositions de "modèles" d'évaluation.

Cette institutionnalisation de l'évaluation est un phénomène propre à ce pays, il ne se rencontre pas en Europe. Monnier (1987) l'analyse comme un double phénomène politique : d'une part, la volonté du gouvernement fédéral de contrôler les moyens financiers mis à la disposition des administrations locales et, d'autre part, la nécessité de justifier sa propre activité face au Congrès.

Le *General Accounting Office* (GAO) est un organisme de contrôle de l'exécutif créé en 1921 par le parlement américain. Le GAO dépasse progressivement sa mission initiale de vérification comptable (paradigme administratif) pour, après la guerre, s'attacher à vérifier l'efficacité globale des politiques mises en œuvre. L'évaluation devient une pratique systématique. D'une moyenne de 200 évaluations par an avec un budget moyen de 100.000 de dollars au début de la décennie, ce secteur passe à 1.315 évaluations en 1977 avec un budget global de 63,6 millions de dollars. Le département de la santé, de l'éducation et du bien-être passe commande pour 38 % de ces évaluations, dépassant de loin les autres départements.

A cette demande croissante répond une véritable industrie de l'évaluation. Deux mille firmes ou sociétés privées fournissant des services contractuels spécialisés en matière d'évaluation sont recensés en 1980. D'autres signes de cette professionnalisation du secteur sont apparents : multiplication des journaux, des réseaux professionnels, des formations et des centres d'études.

Les résultats de ce professionnalisme sont en demi-teinte (Stufflebeam et Shinkfield, 1985) : une meilleure communication, une meilleure formation et une meilleure qualification, plus de coopération avec un niveau d'exigence et de qualité croissant ; mais, par contre, dans la majorité des cas, peu de changement des pratiques de terrain, seules les techniques s'étant diversifiées.

Les paradigmes développés au fil du temps sont tous toujours bien présents.

Ce qui apparaît le plus clairement est le débat très vif entre les positivistes / quantitatifs et les phénoménologistes / qualitatifs. Ce débat, très idéologique au début des années septante, s'est ensuite transformé pour évoluer vers des positions plus pragmatiques cherchant à combiner les avantages réciproques des méthodes. Cependant aucun modèle unificateur ne semble pouvoir rallier les différentes positions et approches. On citera en exemple, pour l'approche positiviste, les travaux de Cook et Campbell en 1979 (*Quasi-experimentation : design and analysis issues for field setting*) ou de Rossi et Freeman en 1982 (*Evaluation : a systematic approach*) ; pour l'approche phénoménologiste les ouvrages de Parlett et Hamilton en 1972 (*Evaluation as illumination : a new approach to the study of innovatory programs*), de Stake en 1975 (*Evaluating the arts in education : a responsive approach*, cité par Stufflebeam et Shinkfield, 1985), de Guba en 1978 (*Toward a methodology of naturalistic inquiry in educational evaluation*). Les tentatives de synthèse peuvent être illustrées par les travaux de Worthen en 1981 (*Journal entries of an eclectic evaluator*). Certains ont proposé d'utiliser l'une ou l'autre approche en fonction des situations d'évaluation rencontrées : la perspective du commanditaire, le caractère routinier ou innovant du programme, les destinataires, ... (Patton, 1981 : *Creative Evaluation*).

A l'approfondissement des méthodologies de la recherche évaluative, on peut associer le mouvement de la méta-évaluation analytique et quantitative (Cook et Gruder, 1978, cité par Straw et Cook, 1990).

La méta-évaluation est surtout développée par les partisans de la recherche évaluative. L'objectif de la méta-évaluation est d'identifier les biais potentiels de tout type que l'on peut rencontrer dans une évaluation et d'utiliser une série de méthodes statistiques ou autres pour en estimer l'importance. C'est une méthode de recherche où les données de multiples études primaires sont réanalysées. L'analyse de multiples études indépendantes permet de rechercher une convergence qui indique une généralisabilité plus large et augmente ainsi la validité des conclusions sur l'efficacité d'un programme.

3. Le paradigme endosystémique, phénoménologique.

C'est au début du véritable développement de l'évaluation et de sa professionnalisation qu'apparaît ce troisième paradigme (Monnier, 1987) que nous associons à l'évaluation explicite formelle.

C'est une pensée du complexe où l'investigation d'un phénomène passe par une compréhension du contexte naturel et du système de repères de l'observateur. Il fait une place importante à l'analyse des aléas, des désordres, paradoxes et incertitudes. L'observateur est l'outil d'émergence de la connaissance.

Pour le "naturaliste", la réalité doit s'appréhender dans sa globalité. Les éléments d'un système ne sont pas autonomes, mais interdépendants. Les résultats ont une valeur relative à l'environnement dans lesquels ils ont été établis. Il n'y a pas de connaissance objective en soi. Les hypothèses varient au cours du processus.

Les évaluations naturalistes ont une tradition moins ancienne que les approches expérimentales ou psychométriques. Elles apparaissent et se développent à partir des années septante. De nombreux modèles ont intégré cette approche pour faire face à la nécessité d'une plus grande flexibilité, d'une attention aux processus, aux besoins de divers publics, aux multiples critères de jugement des composantes d'un programme. L'évaluation naturaliste apporte des descriptions riches et détaillées et des explications globales qui tiennent compte de la complexité du "monde réel".

C'est en décembre 1972 que se tient la "*Churchill College Conférence*" à Cambridge (Parlett, 1974) qui lance le mouvement de l'évaluation "illuminante". Elle réunit des évaluateurs comme Stake, Hamilton, Parlett et MacDonald et des décideurs d'organismes nationaux et internationaux. Le but de la conférence est d'essayer de développer des alternatives radicales, crédibles et viables au paradigme traditionnel. Les conclusions de cette conférence seront diffusées sous la forme d'un manifeste.

L'évaluation naturaliste est une évaluation dans le sens où il s'agit d'une recherche d'information systématique pour établir un jugement sur la valeur d'une innovation, d'un programme, d'un produit. Elle est de type formel parce que, comme l'approche expérimentale, elle s'appuie sur une méthode, un processus en s'inspirant du champ de la sociologie ou de l'anthropologie. L'évaluation "expérimentale" ne se définit que par sa méthodologie. Pour certains, la mesure est l'essence de l'évaluation. Quant à l'évaluation "endosystémique", si elle propose une autre démarche méthodologique, elle va aussi s'intéresser aux autres composantes de l'évaluation.

4. La primauté du paradigme expérimentaliste.

Les trois paradigmes dont nous venons de suivre l'émergence et le développement mettent tous en évidence la composante méthodologique de l'évaluation. Ils seront le point de départ d'autres propositions rendant les deux autres composantes plus explicites.

Néanmoins le paradigme expérimentaliste reste de loin le plus utilisé. Serait-ce parce qu'il est supérieur aux autres ? Certains évaluateurs avancent d'autres hypothèses (Alkin, 1990, Monnier, 1987).

D'une part, pour les utilisateurs de l'évaluation, le discours scientifique est le modèle du discours vrai et la démarche scientifique est considérée comme totalement dévouée à la recherche de la vérité. Il n'est pas étonnant que le paradigme expérimentaliste en évaluation soit le plus demandé par les autorités politiques et administratives. C'est lui qui possède la plus forte crédibilité d'apparence aux yeux de celles-ci et du public. Sa valeur symbolique est la plus forte. La proximité méthodologique entre la recherche et l'évaluation entraîne de nombreuses confusions entre ces deux démarches.

D'autre part, pour les producteurs d'évaluation qui travaillent dans les milieux scientifiques, le contexte académique impose le plus souvent ce modèle comme critère de reconnaissance de la qualité. Les évaluateurs jouant la sécurité, le modèle de la recherche expérimentale conserve la tradition la plus forte.

Stufflebeam et Webster (1980) situent les tests standardisés (paradigme statistique) et les études de recherche expérimentale (paradigme expérimental) dans la catégorie des quasi-évaluations parce qu'ils ne peuvent mesurer réellement la valeur d'un programme.

Les études par tests standardisés comparent les performances à des normes établies. Ces démarches apportent des informations fidèles et valides sur la performance des étudiants, mais ne sont que des indicateurs incomplets d'un programme.

Les études de recherche expérimentale vérifient les effets d'une intervention sur les résultats finaux. Ces études sont de bonne qualité méthodologique, mais leur objet est souvent limité aux intérêts du chercheur, par conséquent, elles sont peu applicables aux situations de terrain.

Dans sa taxonomie, House (1978) associe l'analyse de système sur base économique à une évaluation dont l'objectif est utilitaire et l'épistémologie objectiviste. Les évaluations de coûts et d'efficacité font partie des évaluations formelles. House (1986) fait aussi remarquer le lien entre le modèle de la production industrielle et l'évaluation "scientifique" des programmes sociaux comme celle défendue par Rossi et Freeman (*Evaluation : a systematic Approach, 1979*). Les métaphores sont les mêmes : entrées, sorties, parcours de production, traitement, temps, coûts, procédures, contrôle du processus, qualité du produit final. House en conclut que les modèles d'évaluation basés sur l'approche expérimentale, statistique, psychométrique sont déterminés par les modèles sociaux tels que l'empirisme et le libéralisme.

Alkin et Ellett (1990) attirent notre attention sur la relation entre valeur d'un jugement et évaluation "scientifique". Ce type d'évaluation n'est pas indifférent à la valeur des conclusions d'une étude, mais ils considèrent que cette valeur est garantie par le processus. Pour beaucoup d'ailleurs, ces conclusions ne sont pas un jugement, mais des faits.

Pour la plupart des évaluateurs actuels (Walberg et Haertel, 1990), l'importance du jugement et de la décision est réelle, mais la qualité de la recherche (processus) est un prérequis vital à l'évaluation.

III.C. LES MODALITES DE L'EVALUATION FORMELLE

Les 4 modalités de l'évaluation formelle

	INTERNE			EXTERNE		
INFORMEL	Pii	J	D	Pei	J	D
FORMEL	Pif	J	D	Pef	J	D

La confrontation des deux modalités d'évaluation explicite formelle (Pif, Pef) aux paradigmes exposés met en évidence l'absence de considération pour les démarches méthodologiques menées au niveau interne.

Selon les professionnels de l'évaluation (Alkin, 1990), l'évaluation menée par un praticien (Pif : formelle interne) est possible. Mais à quoi sert-elle ? Pour certains, le praticien n'a pas besoin d'évaluation sur son travail. Il en a une connaissance intime suffisante et supérieure à tout évaluateur externe.

Pour d'autres, ce type d'évaluation peut servir à modifier ou à améliorer un programme (composante "décision"). Tous s'accordent sur un point. L'évaluation interne ne peut permettre de juger et de justifier la qualité d'un programme (composante "jugement"). Aucun ne prend position quant à la validité méthodologique de cette évaluation interne, qui ne suit pas les canons de la science.

Weiss (1986) nous fait part de certaines expériences originales menées aux Etats-Unis. Des groupes de quartier ont été financés pour faire leurs propres évaluations selon leurs perspectives, leurs intérêts, leurs besoins en informations et leurs critères de succès. Ces initiatives auraient été peu satisfaisantes, selon Weiss. Les évaluations intéressent peu les acteurs. Le processus d'évaluation est long, laborieux et ne semble pas leur apporter plus que l'expérience directe. Ils sont intéressés à changer, à agir et pas à étudier. De plus, la qualité méthodologique des évaluations locales est jugée faible et inconstante.

L'évaluation interne jouit donc de peu de considération chez les professionnels et la raison en est souvent méthodologique puisque, pour une majorité d'entre eux, la valeur des conclusions dépend de la rigueur de la démarche.

L'analyse du processus est un point clé de l'évaluation. Cependant, plutôt que d'examiner la qualité méthodologique formelle d'une évaluation interne en prenant comme référence le paradigme expérimental, il a semblé plus utile de décrire la présence ou non d'une énonciation et / ou d'une formalisation. Nous faisons l'hypothèse que ces deux indicateurs seront plus significatifs du rôle de l'évaluation dans le cycle de l'action que le degré de formalisation du processus ou la nature de celui-ci (type de schéma, données qualitatives ou quantitatives, etc.). Nous observerons donc si l'acteur d'un programme signale l'existence d'un processus (énonciation) et s'il en précise le contenu (objets et instruments d'évaluation, mode d'analyse, ...).

Mais d'autre part, nous pourrons aussi examiner le degré de formalisation du processus des évaluations externes. Lorsqu'il y a énonciation d'un processus dans une évaluation externe, ce processus est-il toujours formalisé ?

IV. L'ÉVALUATION DE CONTRÔLE

T4 : L'évaluation de contrôle (jugement et décision explicites)

Jugement de valeur	Décision	<i>Explicite</i>
Processus		<i>Implicite</i>

Les formes les plus anciennes d'évaluation instituées sont signalées par Barbier (1990). Il les met en relation avec l'existence d'un corps de fonctionnaires, en donnant l'exemple du système d'évaluation des personnes dans le système mandarin chinois. Le rôle croissant des administrations accompagnant la naissance des systèmes sociaux au 19^{ème} siècle expliquerait également le passage d'un système d'évaluation de personnes à un système d'évaluation d'actions. Madaus, Stufflebeam et Scriven (1983) notent l'existence d'un système de contrôle administratif au 19^{ème} siècle en Grande-Bretagne dans les secteurs scolaires, sociaux et sanitaires. Le système d'inspection administratif est une forme d'évaluation de contrôle qui naît avec les états.

IV.A. LE PARADIGME JURIDIQUE

Monnier (1987, pp. 74 à 78) trace les grandes lignes du paradigme "juridique" de gestion des affaires publiques.

Dans ce paradigme, il s'agit de vérifier la concordance par rapport à des règles juridiques formelles, des lois qui garantissent l'intérêt public. Le contrôle administratif se cantonne le plus souvent à la vérification du respect des règles. C'est la mission d'inspection.

L'évaluation bureaucratique n'a donc pas pour objet l'amélioration, la pertinence ou l'efficacité, mais le respect des règlements. C'est une vérification a posteriori de l'action aux moyens de procédures définies préalablement. C'est une forme de sanction postérieure au déroulement des opérations dans le cadre de rapports hiérarchiques. Ce type d'approche administrative repose sur une vision simpliste de la complexité de l'évolution des problèmes sociaux et de la complexité de l'administration comme organisation ; il est donc incapable de rendre compte des phénomènes sociaux et d'y intervenir.

C'est néanmoins une pratique d'évaluation dans la mesure où elle vise à attribuer une valeur à une action par rapport à des normes fixées pour aboutir à une décision. C'est une pratique socialement instituée, car les jugements, les normes, les décisions (le plus souvent des sanctions) sont énoncés, observables.

IV.B. LE PARADIGME TECHNICO-ECONOMIQUE

Un autre modèle administratif émerge alors (Monnier, 1987, pp. 78 à 80) : le paradigme technico-économique de gestion des problèmes sociaux.

L'assimilation de la gestion des affaires publiques à celle de l'entreprise privée ne modifie que peu le modèle précédent. La primauté des règles juridiques continue de prévaloir, mais les règles de fonctionnement perdent en clarté. L'administration va intervenir sur la base d'une légitimité intrinsèque de nature technique et économique, mais sans interagir davantage avec les problèmes sociaux et leur nécessaire négociation. L'élargissement du contrôle aux aspects financiers et techniques de l'utilisation des moyens ne change guère le modèle. Même la création d'institutions para-administratives censées offrir plus de souplesse dans la prise de décision et l'action a souvent été un échec. Soit la souplesse n'était pas au rendez-vous à cause d'une tutelle trop présente. Soit la clarté du fonctionnement était absente avec une perte d'indépendance vis-à-vis des influences socio-politiques, la prise de conscience de ces influences pouvant s'accompagner d'une perte de confiance des acteurs sociaux.

Ces considérations de Monnier permettent de saisir la relation entre les composantes "décision" et "jugement" dans l'évaluation de contrôle.

IV.C. LES LIENS AVEC L'EVALUATION D'EXPERTISE

Nous avons vu comment un système de gestion, mis en place pour planifier et diriger, pouvait, par la nécessité de contrôler selon des normes explicites propres à une administration publique, mettre davantage l'accent sur l'énonciation du jugement de valeur.

D'autres modèles, que nous avons déjà abordés, peuvent à certaines occasions être assimilés à un type d'évaluation "jugement - décision".

Les commissions parlementaires sont aussi un moyen de contrôle des politiques. Nous l'avons évoqué en relation avec l'expertise axée sur le jugement de valeur. Il est évident que le lien avec la décision politique peut être explicite comme il peut ne pas l'être, puisque les pouvoirs législatif et exécutif sont différents. De même, le processus de construction du jugement de valeur peut lui aussi devenir plus explicite.

L'accréditation, que nous avons aussi située dans l'évaluation d'expertise, peut prendre d'autres visages. En effet, si au départ l'accent est mis sur la constitution de normes par les professionnels, le pouvoir de décision peut accompagner le système de façon explicite soit au niveau interne, par une révision suite aux recommandations des experts, soit au niveau externe quand il se situe dans un système de contrôle public ou privé (par exemple dans le cas des assurances).

A contrario, un mode d'évaluation par inspection peut se situer davantage dans le registre de l'expertise que dans celui du contrôle. Un fonctionnaire viendra vous rendre "visite", affirmera ne pas venir faire d'évaluation (de contrôle) et vous transmettra son rapport sur ce qu'il a pu constater. La sincérité de votre interlocuteur n'est pas en cause, mais l'ambiguïté est bien présente. Ce rapport informel pourrait ou non, selon des éléments conjoncturels, entraîner des procédures plus réglementées et plus lourdes de conséquences, le pouvoir de tirer les conséquences du jugement étant peut-être aux mains d'un autre service ou d'un autre opérateur.

IV.D. LES MODALITES DE L'EVALUATION DE CONTROLE

Le premier axe de différenciation entre les pratiques d'évaluation d'expertise et de contrôle est donc bien le caractère implicite ou explicite des trois composantes de l'évaluation. Cependant, l'introduction de la différenciation interne - externe et informel - formel sur deux composantes explicites (le jugement et la décision) au lieu de une seule multiplie le nombre de modalités potentielles de l'évaluation de contrôle. Nous les parcourons ci-dessous.

Les modalités de l'évaluation de contrôle

Processus implicite	JUGEMENT INTERNE				JUGEMENT EXTERNE					
			Décision interne	Décision externe			Décision interne	Décision externe		
JUGEMENT INFORMEL	P	Jii				P	Jei			
		Déf	Def	Dif	Def					
		Déf	Def	Dif	Def					
JUGEMENT FORMEL	P	Jif				P	Jef			
		Déf	Def	Dif	Def					
		Déf	Def	Dif	Def					

L'axe interne - externe permet de délimiter quatre formes d'évaluation de contrôle.

- La première comprend un jugement de valeur interne et une décision interne. La gestion du programme est interne. La décision et le jugement de valeur qui la fonde sont énoncés dans le cadre d'une situation de communications et d'échanges internes au programme : réunion d'équipe, conseil de direction. Toutefois le recours à l'énonciation favorise la diffusion du jugement et de la décision vers l'extérieur du programme.
- La seconde consiste en un jugement de valeur interne accompagné d'une décision externe au programme. C'est le cas où une décision externe sur le devenir du programme (le financement, par exemple) se base explicitement sur l'énonciation de la valeur du programme dans un rapport rédigé par ses acteurs.
- La troisième concerne une décision interne prise sur la base d'une expertise externe.
- La quatrième est totalement externe : jugement de valeur et décision. C'est le modèle, par excellence, de l'inspection.

L'axe de la formalisation multiplie par quatre le nombre de modalités.

Ces quatre formes d'évaluation de contrôle interne ou externe peuvent à leur tour être différenciées sur l'axe de la formalisation pour nous permettre de distinguer ou non un jugement formalisé à l'aide de critères et de normes et une décision directe / immédiate ou indirecte / différée. Chacun des axes de différenciation nous permet de prédire l'existence de diverses modalités d'évaluation correspondant à des situations particulières.

Sans les passer systématiquement en revue, on pourrait rencontrer, par exemple :

- une décision interne immédiate, précise, établie sur la base d'un jugement énoncé, mais non motivé, sans critères (P Jii Dif) ;
- un rapport interne mettant en valeur un programme sur base de critères précis, mais conduisant à une décision externe (positive ou négative), qui serait énoncée après un laps de temps considérable (P Jif Dei) ;
- une expertise externe intuitive, sans normes, qui serait un élément parmi d'autres d'une décision interne plus globale (P Jei Dii).

V. L'ÉVALUATION ORIENTÉE VERS LES CONCLUSIONS

T5 : L'évaluation orientée vers les conclusions (processus et jugement explicites)

Processus	Jugement de valeur	<i>Explicite</i>
	Décision	<i>Implicite</i>

Le développement des modèles d'évaluation centrés sur les procédures pour émettre un jugement de valeur prend véritablement son essor au milieu des années soixante avec des auteurs comme Scriven, Stake et Cronbach (Stufflebeam et Shinkfield, 1985).

D'un modèle d'évaluation ouvert sur le jugement de valeur et même sur la décision, on retourne vers une approche centrée sur la mesure des performances individuelles.

V.A. SCRIVEN

Scriven (Stufflebeam et Shinkfield, 1985) est un mathématicien et un philosophe des sciences qui enseigne l'évaluation à l'université de Californie entre 1965 et 1975.

Il définit l'évaluation comme la détermination systématique et objective de la valeur ou du mérite d'un objet ou d'une action. L'évaluation est dite objective, car elle utilise les instruments de collecte de données, la pondération et la sélection des buts. C'est une activité méthodologique systématique. Elle est la mieux réalisée par un évaluateur externe et indépendant qui rend un jugement sur un objet en accumulant des comparaisons avec d'autres objets dans la rencontre des besoins des consommateurs.

L'évaluation est donc comparative, orientée vers les besoins des consommateurs, réalisée professionnellement par des procédures systématiques, conduite objectivement et elle aboutit à des jugements et des recommandations.

C'est ainsi que Scriven critique les commissions d'accréditation composées de professionnels qui utilisent des critères intrinsèques (qualification du personnel, livres disponibles, ...) et sont incapables de mesurer la performance réelle. C'est une critique de l'évaluation centrée sur les moyens et sur le jugement professionnel ! La démarche évaluative doit se professionnaliser, mais avec des critères objectifs et explicités.

Deux concepts sont particulièrement attachés au nom de Scriven, celui d'évaluation indépendante des buts (*Goal-Free Evaluation*) et la distinction entre évaluation formative et évaluation sommative.

1. L'évaluation indépendante des buts.

L'évaluation indépendante des buts (Stecher, 1990, Stufflebeam et Shinkfield, 1985) est une approche de l'évaluation dans laquelle la valeur est déterminée par l'examen de tous les effets du programme sans référence aux buts ou aux objectifs de celui-ci.

L'évaluation "*Goal-free*" est une évaluation basée sur les besoins. L'important pour un programme n'est pas de réussir à atteindre des buts fixés, mais de répondre réellement à des besoins démontrés. Le premier travail de l'évaluateur est donc de déterminer les besoins.

Cette focalisation sur les besoins plutôt que sur les intentions est plus un principe philosophique qu'un ensemble de règles et de procédures sur la manière de mener une évaluation. Scriven veut éviter une approche trop étroite et trop orientée par les responsables des programmes. L'évaluateur doit porter un jugement qui soit garant des besoins de la société.

L'évaluateur doit rester ignorant des buts définis par le programme et chercher tous les effets du programme. Cette démarche n'est pas opposée à celle basée sur les objectifs, mais complémentaire. Elle peut être menée avant celle-ci ou en même temps, mais par des évaluateurs différents. Les avantages, selon Scriven, sont une moindre intrusion dans le programme, une moindre sensibilité aux changements d'objectifs et aux biais sociaux, perceptuels et cognitifs ; une meilleure sensibilité aux effets cachés, plus d'équité dans l'examen des valeurs à prendre en compte, plus de professionnalisme.

La difficulté de l'évaluation indépendante des buts est cependant d'assigner une valeur aux conclusions puisqu'il n'y a pas de référence sur ce qui a voulu être fait. Comment distinguer les conséquences désirables ou indésirables ?

Scriven prend pour étalon les besoins des consommateurs. Tout l'enjeu porte alors sur la définition de ces besoins. Scriven propose l'analogie avec une approche consumériste où la valeur d'un produit ne doit pas être établie en fonction du producteur, mais du consommateur. Dans l'approche "orientée vers le consommateur" de Scriven, le rôle de l'évaluateur est d'éclairer le consommateur. Armé des techniques lui permettant d'obtenir des informations pertinentes et exactes, adoptant une vue profondément raisonnée de l'éthique et du bien commun, l'évaluateur aidera les professionnels à produire des produits et des services qui auront une haute qualité et une grande utilité pour les consommateurs. Plus important, l'évaluateur aidera les consommateurs à identifier et à juger les biens et les services alternatifs.

La proposition de Scriven n'a pas véritablement abouti à des démarches pratiques, pour plusieurs raisons : la résistance des décideurs et des gestionnaires de programme à se prêter à un examen aussi large et sans buts clairs ; le refus de nombreux évaluateurs de s'engager dans une responsabilité aussi importante ; mais aussi la difficulté matérielle et financière de mener des études sur tous les effets. Néanmoins, ce modèle a largement attiré l'attention sur l'indépendance de l'évaluateur et sur les effets non attendus ou cachés d'un programme. Il a largement inspiré d'autres auteurs qui ont proposé des démarches méthodologiques plus précises allant dans le même sens.

Des critiques assez vives ont aussi été portées à ce modèle. Les détracteurs font remarquer que l'évaluateur substitue ses finalités et ses objectifs à ceux du programme même s'il se réfère au bien commun. Patton (1978), notamment, reproche à ce type d'évaluation de court-circuiter le débat fondamental sur les objectifs d'une action.

2. L'évaluation formative et sommative.

En 1967, Scriven distingue une fonction formative pour aider au développement des programmes ou d'autres objets et une fonction sommative qui mesure la valeur de l'objet.

L'évaluation sommative sert à décider pour un curriculum déterminé s'il représente une amélioration significative parmi les alternatives disponibles. Elle est réalisée par un évaluateur externe et les conclusions sont rendues publiques. Tous les effets sont recherchés et examinés par rapport aux besoins. Ce type d'évaluation se préoccupe des coûts. La validité des buts en relation avec les besoins y est parfois étudiée, à défaut d'un examen de tous les résultats indépendamment des buts.

L'évaluation formative fait partie du processus de développement pour assister la planification et la production. Elle s'intéresse au contenu, à l'usage, à la viabilité du matériel, etc. Toutefois, l'évaluateur, en voulant aider à améliorer les programmes, se compromet trop souvent et échoue à juger objectivement ce qui est le fondement de l'évaluation. L'évaluation formative ne peut être envisagée que comme un complément.

De façon générale, par la suite, les études à petite échelle sur des aspects particuliers d'un programme au cours de son développement ont été appelées "formatives" et les études comparatives examinant les résultats de programmes terminés, dans le but de faire des recommandations sur leur poursuite ou leur sélection, ont été appelées "sommatives". Dans une première analyse rapide, on y retrouve la distinction entre un type d'évaluation orienté vers le jugement (sommatif) et un type d'évaluation orienté vers la décision (formatif).

Cette distinction entre évaluation formative et sommative sera approfondie dans la présentation du type d'évaluation suivant : "processus - décision".

V.B. STAKE

Les modèles d'évaluation de la "contenance" et d'évaluation répondante de Stake donnent aussi à l'évaluateur un rôle de juge, non pas comme garant de la société, mais comme arbitre entre les valeurs de différents groupes.

En 1967, Stake (Stufflebeam et Shinkfield, 1985) propose d'abord un modèle appelé "*The Countenance of Educational Evaluation*". Ce modèle de la "contenance" est construit sur la même base que le modèle de Tyler avec une comparaison entre des résultats attendus et observés, mais Stake élargit la comparaison aux processus, aux normes et aux jugements.

1. Principes du modèle de la "contenance".

Cette approche ne propose pas un guide pour construire une évaluation, mais des principes généraux pour faciliter cette tâche. Elle veut attirer l'attention sur tous les objets à évaluer. C'est un modèle de la "totalité" s'inspirant du concept de "qualité totale" qui se développe à l'époque dans l'industrie. Ses principales caractéristiques sont :

- a) la distinction au sein de l'évaluation entre une phase descriptive et une phase de jugement ;
- b) l'élargissement des objets d'évaluation aux antécédents et aux transactions en plus des résultats ;
- c) l'analyse comparative entre les intentions et les observations ;
- d) l'identification des normes pertinentes et souvent conflictuelles.

L'évaluateur va recueillir des informations de différentes sources et avec différentes techniques sur une large variété d'objets pour comparer les intentions et les observations sur les antécédents, les transactions et les résultats.

a) La description et le jugement.

En considérant la description comme un acte de base de l'évaluation, Stake suit la tradition de Tyler, mais en diversifiant et en élargissant les objets à décrire.

En tenant compte de la dimension du jugement, Stake s'interroge sur la seule responsabilité de l'évaluateur dans ce domaine. C'est en effet, pour lui, une position irréaliste. Elle va à l'encontre de la position d'observateur neutre que doit tenir l'évaluateur et peut entraîner un rejet du contrôle exercé de cette manière par l'évaluateur. Le compromis proposé par Stake est de collecter objectivement les opinions et jugements des différentes parties : porte-parole de la société, experts, enseignants, parents, étudiants. Il n'adhère donc pas complètement à la recommandation de Scriven de faire de l'évaluateur le juge final. Il n'y a pas de jugement universel, valable pour tous et porté au nom du bien commun par l'évaluateur.

Stake attire aussi l'attention de l'évaluateur sur les différentes rationalités autres que la sienne et qu'il pourrait percevoir comme ambiguës, inconsistantes, illégales ou immorales.

b) Les antécédents, transactions et résultats.

Les antécédents sont les multiples informations relatives aux conditions préalables à l'enseignement et à l'apprentissage et qui sont en relation avec les résultats : l'état de santé de l'enfant, la position du syndicat vis-à-vis de la formation continue, ...

Les transactions sont les innombrables rencontres de l'apprenant avec son milieu d'apprentissage.

Les résultats sont ce qui a été produit par le programme : rendement, attitudes, habiletés, aspirations, ... Ils comprennent les résultats attendus et inattendus.

c) La comparaison.

L'évaluateur compare les intentions aux observations. Il recherche les différences entre ce qui était attendu et ce qui arrive réellement. Il cherche aussi à établir des liens entre les antécédents, les interactions d'enseignement et les résultats.

d) Les normes.

Les normes sont des critères explicites pour mesurer la qualité d'une situation éducative. Face à la complexité de la question des critères, Stake conseille aux évaluateurs d'identifier de la façon la plus large et la plus claire possible quelles sont les normes en regard d'un programme.

L'évaluateur accepte que différentes parties puissent avoir différents critères de jugement et que les critères puissent aussi dépendre du stade de développement d'un programme.

Stake propose donc une démarche systématique et élargie : il reconnaît la place du jugement dans l'évaluation tout en fixant un rôle d'observateur à l'évaluateur.

2. Principes du modèle "répondant".

De son modèle précédent, Stake (Stufflebeam et Shinkfield, 1985) retient l'importance des normes et des valeurs.

L'évaluation de Stake est répondante parce qu'elle est attentive aux multiples normes parfois contradictoires tenues par différents groupes. Son but est de servir les besoins en évaluation des différentes parties intéressées. Elle est aussi répondante, car elle s'adapte continuellement dans ses buts et ses moyens de rassembler de l'information.

C'est une évaluation centrée sur les activités, les sujets d'intérêts plutôt que sur des hypothèses ou des objectifs. L'évaluateur propose une approche pluraliste, flexible et interactive, orientée vers la "clientèle" de l'évaluation. Selon Stake, une étude sera identifiable comme "répondante" si elle révèle comment les sujets d'étude ont été négociés et les publics servis.

Il reconnaît la difficulté d'identification de son approche qui ne prône pas un modèle fixe.

a) L'approche "répondante" est subjectiviste.

Elle part de l'idée que la valeur d'un programme dépend de la perception subjective de cette valeur par ceux qui sont concernés. Elle reconnaît aussi que la subjectivité du chercheur est en jeu dans le choix, la mise en oeuvre et le compte rendu de ses observations. La subjectivité n'est pourtant pas un obstacle à l'évaluation. L'évaluateur conscient de cet enjeu prendra soin d'éclairer cette subjectivité, de la faire apparaître, de veiller à la validité en recoupant ses sources d'information.

b) L'approche "répondante" est pluraliste.

L'évaluateur est au service des diverses personnes qui sont affectées par le programme. Il doit rester cependant lucide et indépendant, même si certaines conclusions ne sont pas favorables à certains partenaires. Il est conscient que les enjeux autour d'une évaluation sont multiples et ne concernent pas seulement la compréhension et la prise de décision (par exemple, la légitimation, la protection contre les critiques), qu'ils cachent même parfois des buts inavoués (par exemple, la discrimination raciale).

c) L'évaluation "répondante" n'est pas par nature une recherche naturaliste ou qualitative.

Toutes les méthodes sont utilisables. Elles sont négociées en fonction des besoins et choisies en dernier ressort par l'évaluateur. Bien que les tests standardisés soient rarement adaptés à des situations particulières et que leur mise au point soit coûteuse, ils ne sont pas exclus, mais jouent un rôle accessoire. Les personnes sont plus des sources d'information que des sujets d'étude. Elles sont questionnées sur leur vision du changement et pas pour savoir comment elles ont changé. Les ressources étant limitées, l'évaluateur envisagera très tôt les publics à informer et les modalités de ses rapports. Ils seront attractifs et compréhensibles, adaptés aux demandes. Pour certains, il utilisera des portraits et des témoignages verbaux ; pour d'autres, des données et des traitements statistiques. Il y aura également souvent des rapports intermédiaires et des mises au point sur la poursuite du travail d'évaluation.

Les deux modèles de Stake montrent bien toutes les nuances que l'on peut découvrir dans des modèles qui fondamentalement mettent en œuvre le même schéma : l'importance accordée à la fois au processus d'évaluation et au jugement de valeur qu'il produit. Tout d'abord, Stake dépasse le modèle de Tyler en restant attaché à une approche méthodologique assez rigide, mais en intégrant déjà le pluralisme des valeurs. Ensuite, sa position méthodologique va se diversifier pour mieux répondre à ce pluralisme, en intégrant les apports des modèles "naturalistes" comme ceux de Guba ou de Palett et Hamilton.

Découvrons un dernier auteur, de la même époque, qui propose encore une autre vision de la relation entre le processus et le jugement de valeur.

V.C. CRONBACH

Dès 1963, Cronbach (Stufflebeam et Shinkfield, 1985) indiquait l'importance de l'évaluation comme moyen de développer et d'améliorer l'éducation.

1. L'évaluation et l'évaluateur font partie de l'action politique.

Au centre de la réflexion de Cronbach se situe clairement l'affirmation du caractère politique de l'évaluation. L'évaluateur a une responsabilité sociale, mais il est soumis aussi à toutes les pressions sociales.

Les préoccupations politiques sont à l'origine des demandes d'évaluations. Elles influencent le choix des questions explorées et le mode de récolte des informations. Elles déterminent l'utilisation des conclusions. Mais tout en identifiant l'importance de la décision sur l'activité évaluative, Cronbach reste attaché à un type d'évaluation centré surtout sur le processus bien qu'il intègre le jugement. Comme pour les tenants de l'évaluation "formelle", la valeur d'une évaluation est déterminée par sa qualité scientifique, bien qu'elle doive être aussi significative pour la communauté politique qu'il définit comme les clients, l'équipe du programme, l'administration, les citoyens intéressés.

2. L'évaluation doit être compréhensive plus qu'explicative.

L'effort de l'évaluateur portera principalement sur la conception (*designing*) de son évaluation pour la rendre appropriée à chaque situation. Cronbach ne croit pas à la supériorité d'un type de méthodologie sur l'autre et prône un mélange des styles en fonction des situations pour atteindre la meilleure compréhension possible. Il privilégie pourtant les méthodes les plus structurées.

Il n'y a pas de conflit entre le contrôle expérimental et l'information qualitative, entre une approche humaniste et scientifique. Il conseille cependant de préserver une certaine homogénéité dans une même étude.

Le critère de choix de l'approche est défini par la meilleure résolution d'un problème politique donné à un moment donné. La réalisation d'une évaluation est un processus réfléchi et évolutif et pas une démarche figée et rationnelle. Il faut y investir du temps et de l'énergie.

3. La communication des résultats de l'évaluation est une tâche importante et délicate. Elle est réalisée vers l'ensemble des participants.

Cronbach insiste aussi beaucoup sur l'importance de la communication entre l'évaluateur et ses publics. La communication peut être informelle, continue. L'évaluateur agit comme un journaliste qui enquête et attire l'attention. Il peut utiliser des incidents révélateurs, des narrations.

L'évaluateur a une responsabilité vis-à-vis d'une communauté large. L'évaluateur a le droit et le devoir d'extrapoler, de fournir des prédictions alternatives. L'évaluateur doit se préoccuper surtout de rapporter ce qui sera utile à la compréhension de ce qui s'est passé et à la définition des orientations futures.

4. L'évaluation doit guider et éclairer plus que juger, car il n'y a pas de vérité universelle.

L'évaluateur est en face de plusieurs types de décideurs. Son but est de répondre à ceux-ci, de réduire les alternatives, de clarifier les sujets de débat, mais aussi de faire avancer la compréhension sociétale sur un type de problème. L'évaluateur est jugé sur sa communication comme un éducateur. Il doit soulever les questions, les élucider, générer des idées. Mais l'évaluateur ne substitue pas son jugement à celui de ses interlocuteurs. Il doit les comprendre et les convaincre de sa démarche.

5. L'évaluation doit éclairer l'action sociale, mais par l'ensemble de ses interventions et non vis-à-vis d'un programme en particulier.

Cronbach fonde beaucoup d'espoirs dans le rôle de l'évaluateur dans la société. Il attire l'attention sur les limites de l'évaluation pour résoudre les conflits de nature politique. Cependant la valeur d'une évaluation ne se limite pas à son usage immédiat, mais à ce que la société peut en retirer comme expérience. L'évaluation ne mesure pas simplement les résultats, mais étudie la valeur des succès et des échecs.

Par bien des côtés, la position de Cronbach et de ses collègues est très proche de celle de Scriven : la mission sociale de l'évaluateur, sa professionnalisation. Elle se rapproche des options de Stake par une vision pluraliste et par une méthodologie moins expérimentaliste, plus pragmatique, plus éclairante. Elle s'écarte néanmoins de la position de Stake comme de la vision "utilitariste" de Patton ou Alkin que nous examinerons plus loin, car elle attache peu d'importance à l'utilisation immédiate et directe des résultats de l'évaluation.

Cet examen un peu plus détaillé de quelques auteurs reconnus dans le domaine de l'évaluation en éducation montre le besoin pour ceux-ci de situer leur travail, leurs pratiques. Les pratiques ne sont plus uniquement définies sur un plan méthodologique, mais en relation avec les buts fixés à l'évaluation. Ces buts sont aussi fixés par rapport au rôle et aux responsabilités des divers acteurs en présence : l'évaluateur, les responsables du programme, ceux qui y sont impliqués, la société. De plus, la décision devient un point de référence même si elle reste en filigrane, essentiellement à cause de la position professionnelle des évaluateurs.

V.D. LES IDEOLOGIES DE L'EVALUATION

La position de Scriven vis-à-vis de ce qu'il appelle les idéologies de l'évaluation permet de situer les modèles et concepts des différents auteurs par rapport à la dimension "valeur" de l'évaluation. En effet, Scriven a classé les approches de l'évaluation en quatre groupes qu'il a analysés et critiqués en profondeur, appelant les évaluateurs à les identifier et à en reconnaître les biais. La base de la réflexion de Scriven se situe dans une analyse de la position de l'évaluateur vis-à-vis de la dimension "valeur" de l'évaluation.

1. L'idéologie séparatiste.

Ce type d'évaluateur se veut indépendant de l'objet évalué. Il refuse de concevoir l'évaluation comme un système établissant ses propres critères (*self-referent activity*). L'indépendance est essentielle pour éviter les biais. L'évaluateur "séparatiste" évite de poser la question du jugement. Sans être séparatiste puisqu'il reconnaît l'existence du jugement, Cronbach pense que l'évaluateur doit rester indépendant de celui-ci.

2. L'idéologie positiviste.

Cette vision de l'évaluation rejette aussi la dimension du jugement dans l'évaluation comme dans la science. Elle se base sur une approche scientifique de l'objectivité et ne cherche pas à assigner une valeur aux conclusions obtenues. L'évaluation serait une technique évitant la détermination des valeurs. En fait, c'est la démarche scientifique expérimentale qui crée la légitimité (la valeur).

3. L'idéologie managériale.

L'évaluateur qui adhère à cette idéologie n'évite pas la question des valeurs, mais il base son travail sur les souhaits des gestionnaires de programme. Selon Scriven, cette position conduit à une impasse. En effet, aider les managers à juger de leurs intentions ne contribue pas à fournir une évaluation véritable au service de la société. Pour juger des résultats, des bénéfiques, l'évaluateur doit analyser les besoins, ne pas se préoccuper des buts prédéfinis, comparer les alternatives, examiner le rapport coût-efficacité, évaluer les personnes comme les programmes.

4. L'idéologie relativiste.

Cette approche est une réaction à la vision positiviste d'une réalité objective mais mesurable. Elle affirme qu'il n'y a pas de vérité objective, que tout est relatif, qu'il y a de multiples perspectives, critères, mesures et réponses, que, trop souvent, la position positiviste ne fait que défendre la position étroite des groupes qui possèdent le pouvoir.

L'idéologie relativiste est illustrée dans la position de Stake (cité par Stufflebeam et Shinkfield, 1985) qui propose d'identifier toutes les valeurs en présence. Certains pensent qu'il est possible de dépasser le pluralisme des valeurs. House, en s'appuyant sur la philosophie sociale de Rawls, propose comme groupe de référence le groupe le moins avantagé de la société.

5. La position idéologique de Scriven.

Cette analyse de Scriven met particulièrement en évidence sa propre position idéologique. Il assigne à l'évaluateur une "mission" sociale. L'évaluateur doit rechercher la rigueur scientifique, l'objectivité, l'indépendance, le bien de la société, mais aussi assumer le jugement de valeur en déterminant les besoins de la société. Scriven reconnaît la pluralité, mais n'accepte pas d'accorder la même valeur à tous les points de vue. Il reste attaché à la découverte d'une réalité objective même si elle est complexe.

En suivant notre proposition, on voit que Scriven entend bien dépasser la vision d'une évaluation de type formel, mais ne veut pas entendre parler non plus d'une évaluation plus sensible au problème de la décision.

V.E. LES MODALITES DE L'EVALUATION ORIENTEE VERS LES CONCLUSIONS

Avant de situer les débats, les modèles et les pratiques sur le terrain de la décision, parcourons les différentes modalités possibles en matière d'évaluation orientée vers le jugement, les conclusions.

Les 16 modalités de l'évaluation orientée vers les conclusions

Décision implicite	JUGEMENT INTERNE				JUGEMENT EXTERNE			
	Processus interne	Processus externe			Processus interne	Processus externe		
Processus informel	Pii	Pei	Jii		Pii	Pei	Jei	
Processus formel	Pif	Pef			Pif	Pef		
JUGEMENT INFORMEL	D				D			
Processus informel	Pii	Pei	Jif		Pii	Pei	Jef	
Processus formel	Pif	Pef			Pif	Pef		
JUGEMENT FORMEL	D				D			

L'axe interne - externe va ici aussi déterminer quatre formes d'évaluation, elles-mêmes réparties chaque fois en quatre modalités selon l'axe informel - formel.

1. La première forme est interne au niveau du processus et du jugement.

Les acteurs directs du programme proposent un processus d'évaluation simplement énoncé (informel) ou construisent une démarche structurée sur le plan méthodologique (formel). Ce processus informel ou formel conduit à des conclusions, à un jugement de valeur, qui est exprimé, communiqué. Les critères de ce jugement aussi ont été précisés ou sont restés informels. La décision reste, elle, implicite.

Par exemple, un membre de l'équipe de gestion du programme est chargé de faire rapidement le diagnostic d'un problème survenu en cours de route. Il fait le point lors d'une réunion en présentant sa démarche et ses propositions, qui sont notées dans le compte rendu. Les conséquences de cette analyse ne sont pas discutées. Chacun est responsable d'adapter son travail (Pii Jii D).

Ou encore, une équipe spécialisée en évaluation appartenant à l'institution responsable du programme met en oeuvre une évaluation d'efficacité dont les conclusions sont intégrées dans le rapport final remis au commanditaire (Pif Jif D).

2. La seconde forme est interne au niveau du processus et externe au niveau du jugement.

Par exemple, un processus d'évaluation interne formel est mis sur pied avec l'aide d'un consultant, les résultats et les conclusions sont soumis à l'analyse critique de cet expert externe, qui remet un avis non formalisé sur base de sa propre expérience (Pif Jei D).

3. La troisième forme est externe au niveau du processus et interne au niveau du jugement.

Par exemple, un évaluateur externe est chargé de réaliser une évaluation formelle en interaction étroite avec les responsables d'un programme. Une liste d'indicateurs et de critères potentiels est préparée par ce spécialiste, mais le choix est réalisé par les acteurs. Le processus formel de recueil des données est conçu et réalisé par cet évaluateur, qui fournit ensuite les résultats de son analyse laissant à l'équipe interne l'initiative d'interpréter les données et d'en tirer les conclusions. Cette démarche fait l'objet d'un rapport qui ne précise pas l'utilisation des conclusions tirées (Pef Jif D).

Le modèle de Cronbach se rapproche de cette situation où l'évaluateur joue un rôle de collecteur d'informations pour divers publics leur laissant la responsabilité des conclusions.

4. La quatrième et dernière forme prévue est entièrement externe.

Nous rencontrons là les modèles de Scriven et de Stake que nous avons développés plus haut. Le processus y est explicite, externe et formel.

Le jugement est clairement externe et formel chez Scriven où l'évaluateur est seul juge de la valeur du programme (Pef Jef D). Ce jugement est aussi extérieur chez Stake même s'il est établi à la suite d'un partage. L'évaluateur joue aussi un rôle de juge-arbitre extérieur.

On pourrait aussi concevoir une démarche totalement externe et non formalisée. Par exemple, l'évaluateur annonce qu'il réalisera une évaluation structurée, mais ne prend pas la peine de détailler son mode opératoire. Les critères de jugement restent également dans le vague (Pei Jei D).

VI. L'EVALUATION "PROCESSUS - DECISION"

T6 : L'évaluation "processus - décision" (processus et décision explicites)

Processus	Décision	<i>Explicite</i>
Jugement de valeur		<i>Implicite</i>

Dans leur proposition de classification des modèles d'évaluation, Alkin et Ellett (1990) mettent en évidence la dimension de l'utilisation. Alkin (Alkin, Daillak et White, 1979) fut un des premiers à étudier de manière approfondie l'utilisation de l'évaluation. Nous aborderons ses recherches dans le troisième chapitre. Cette dimension de l'utilisation introduit la composante "décision" de l'évaluation.

Pour Alkin et Ellett (1990), les modèles proposés se différencient selon la nature et le rôle des utilisateurs de l'évaluation, le degré d'interaction entre l'évaluateur et les utilisateurs. L'approche orientée vers les décideurs met l'évaluation au service de ceux-ci à travers la détermination des questions utiles aux programmes.

Dans le premier modèle de ce type (Stufflebeam et al., 1971), l'évaluateur délimite son travail avec le décideur, puis récolte et fournit les informations. Le modèle de décision est de nature rationnelle.

Dans les modèles ultérieurs, comme ceux proposés par Alkin (1979) ou Patton (1978), la prise de décision est vue comme un processus plus complexe de nature incrémentale où l'évaluateur interagit avec les utilisateurs de manière constante.

Cette composante de décision n'est pas indépendante des deux autres. Alkin et Ellett (1990) font le lien avec les dimensions "valeur" et "méthodologie" qu'ils ont identifiées dans les modèles d'évaluation.

L'approche "décideur / utilisateur" voit l'évaluateur comme un conseiller à la décision auprès de certains publics. Ces modèles considèrent que la valeur intrinsèque d'un programme est déterminée par ce que pensent ces groupes sur ce qui a ou non une valeur. Ceux qui ont la légitimité de décider, ont aussi celle de juger.

L'intégration de la composante "décision" et son lien avec le jugement de valeur permet bien d'identifier des différences fondamentales entre certains modèles. Mais l'identification précise du type d'évaluation que nous désignons comme orienté vers la décision n'est pas aisé. Existe-t-il une évaluation qui n'aurait pas besoin du jugement de valeur et qui établirait un lien direct entre le diagnostic d'un problème et sa résolution dans un cycle de régulation de l'action ? Nous faisons l'hypothèse que non, mais que ce jugement, tout en étant présent, reste implicite. Le passage par l'énonciation du jugement de valeur comme produit du processus d'évaluation serait fonction du contexte de production et d'utilisation de l'évaluation.

VI.A. LES DISTINCTIONS ENTRE L'ÉVALUATION FORMATIVE ET SOMMATIVE

Le débat sur la distinction entre l'évaluation formative et sommative permet d'illustrer cette difficulté. Il n'y a pas de différence de nature entre une évaluation formative et sommative, mais une différence de contexte.

Lewy (1990) offre un rapide aperçu de l'évolution des termes "formatif" et "sommatif".

L'introduction de ces termes par Scriven et la distinction qu'ils sous-tendent a paradoxalement permis de diffuser l'idée qu'il était permis et même hautement désirable de conduire des études d'évaluation pendant le déroulement d'un programme en ne suivant pas les schémas classiques des études comparatives. La conviction apparaît de la possibilité pour l'évaluateur de contribuer à l'amélioration d'un programme.

La distinction formatif - sommatif a donc permis de focaliser et de stimuler le débat sur les fonctions de l'évaluation, renversant même la tendance et conduisant de nombreux évaluateurs à négliger les aspects sommatifs. Cronbach, par exemple, affirme en 1980 que l'évaluation formative est la plus utile et contribue le plus à l'amélioration de l'éducation. L'amélioration d'un programme passe par la disponibilité d'informations sur son déroulement. D'un autre côté, les partisans de l'évaluation comparative reprochent à l'évaluation formative d'être moins rigoureuse sur le plan méthodologique et de mener souvent à des travaux sans intérêts.

Cependant la distinction entre évaluation formative et sommative n'est pas toujours claire.

Les termes "formatif" et "sommatif" et la dichotomie qu'ils recouvrent (amélioration, régulation d'une part et jugement de valeur d'autre part) ont donc fait couler beaucoup d'encre et une signification précise est devenue de plus en plus difficile à trouver. En effet, des jugements de valeur peuvent être portés pendant le déroulement d'un programme et des améliorations apportées au terme de celui-ci. Tout dépend de la taille et de la durée du programme et de la manière dont il s'insère dans une politique plus large.

De nombreux auteurs ont cherché à distinguer ces termes au moyen d'autres dichotomies : prospectif / rétrospectif, évolutif / fixé, naturel / expérimental, holistique / analytique, processus / résultat, etc. Ces distinctions ne sont pas toujours plus éclairantes.

- Une évaluation dite "formative" peut être composée d'une succession d'évaluations de résultats intermédiaires sur des unités d'actions constitutives du programme. Un retour rapide d'informations (feedback) sur le succès ou l'échec permet ainsi de modifier rapidement le programme. Cette évaluation formative peut aussi tenter de poser le diagnostic du fonctionnement du programme pour l'améliorer.
- Le jugement global au terme du programme peut utiliser une méthodologie "naturelle". Il peut tenir compte du processus. Il est rétrospectif, mais permet de fixer de nouveaux buts, de nouvelles stratégies de manière prospective.

A la suite à ces polémiques, Scriven a précisé ces termes (1980).

- Une évaluation "formative" est conduite pendant le développement ou l'amélioration d'un programme, d'un produit ou d'une personne. C'est une évaluation menée pour les gestionnaires du programme, qui reste interne à celui-ci, que l'évaluateur soit interne ou externe.
- Une évaluation "sommative" est conduite après l'arrêt du programme, au bénéfice d'une audience externe ou des décideurs (commanditaires, futurs utilisateurs). Elle peut aussi être menée par un évaluateur interne ou externe.

Pour Scriven, il n'y a donc pas de différences méthodologiques entre les deux évaluations, mais seulement un moment et un public différents. Une même étude peut être vue comme formative ou sommative selon l'utilisateur.

Nous pensons que ces différences de moment et de public sont contextuelles et qu'elles entraînent une énonciation différente du jugement de valeur et de la décision, une position stratégique différente de l'évaluateur et des degrés de formalisation variables. Les évaluations formative et sommative ne sont que des facettes différentes d'un même processus, ce qui rend la distinction difficile sur le plan théorique.

VI.B. LES DIFFERENTS MODELES THEORIQUES ET L'EVALUATION ORIENTEE VERS LA DECISION

Pour Scheerens (1990), l'idée que l'évaluation supporte la décision administrative ou politique peut être considérée comme une caractéristique de base. De même, dans notre système d'analyse, nous avons considéré la décision comme une composante de base de l'évaluation. Scheerens analyse toutes les approches à la lumière de ce postulat. Cela apparaît, pour lui, dans les textes classiques de Suchman et Campbell, défenseurs d'une évaluation scientifique, comme dans ceux de Stufflebeam, Alkin ou Patton, promoteurs de l'évaluation orientée vers la décision ou centrée sur l'utilisation.

La différence entre les approches dépend en réalité du "contexte décisionnel". Scheerens nous propose donc de les réexaminer en fonction d'un modèle "rationnel" ou "cumulatif" de la décision, qui recouvre les caractéristiques suivantes :

Modèle rationnel

Buts clairs

Ils sont définis au préalable et de préférence de manière opérationnelle.

Distinction entre buts et moyens

Les moyens sont définis après les buts.

Hypothèses de causalité

Chaque facteur peut être identifié et conduit aux fins.

Perspective à long terme

La planification est possible.

Programmes stables

Il y a une vue optimiste de l'habileté et de la volonté d'implanter un programme ; l'invariabilité permet l'expérimentation.

Décideurs identifiables

Ils sont autorisés à décider.

Analyse causale

Il y a mesure des performances (produit) et schémas de recherche.

Évaluateur indépendant

L'évaluation est une activité rationnelle, impartiale, répondant à des normes professionnelles.

Études à large échelle

Le programme est planifié et à long terme.

Utilisation naturelle

Elle découle de l'évaluation de manière unique dans une séquence linéaire.

Modèle cumulatif

Buts ambigus

Cette ambiguïté permet la coexistence d'intérêts différents.

Non distinction entre buts et moyens

Il n'y a pas de priorité.

Contexte historique

L'évolution est lente, cumulative et localisée.

Perspective à court terme

Il y a méfiance vis-à-vis de la planification.

Programmes flous et passagers

L'implantation est le problème majeur.

Décision diffuse

La responsabilité est difficile à déterminer ou partagée.

Analyse descriptive

Ce sont des études de cas d'implantation locale, des données qualitatives.

Évaluateur dépendant

L'évaluation fait partie du jeu politique même si l'évaluateur a des intérêts distincts.

Études à petite échelle

On privilégie la diversité des approches.

Utilisation graduelle et progressive

Elle participe à une décision qui se construit pas à pas.

Ces caractéristiques sont plutôt des extrêmes sur un continuum qui admet de multiples variantes.

A la lumière de ces deux modèles de la prise de décision, Scheerens présente plus particulièrement quatre approches de l'évaluation.

Les deux premières, l'évaluation centrée sur l'utilisation et l'évaluation basée sur les acteurs, suivent le modèle de la décision cumulative. Pour illustrer celui-ci, nous avons choisi de présenter le modèle de Patton et le modèle des "acteurs" développé par une agence gouvernementale américaine.

Les deux suivantes, l'évaluation sur le modèle des paris et l'évaluation "reconstruite rationnellement", essaient de restaurer un modèle rationnel. En effet, Scheerens fait partie des auteurs actuels qui tentent de tenir compte des acquis des approches pluralistes, mais qui souhaitent restaurer un modèle scientifique. Scheerens plaide pour maintenir le prérequis vital à l'évaluation qu'est la qualité de la recherche, mais en tenant compte du contexte décisionnel.

Il faut rassembler des informations sur la décision parce le contexte de l'évaluation doit être défini dès la première étape de faisabilité pour trouver le meilleur équilibre entre les aspirations de l'évaluation et les intrications contextuelles d'une situation donnée : degré de coopération possible avec les décideurs, autonomie des acteurs, indépendance du chercheur, ressources disponibles, etc.... Mais, il faut garder la rigueur de la démarche scientifique parce que les données de l'évaluation restent indispensables pour établir le succès d'un programme.

Nous présenterons tout d'abord le modèle d'évaluation orientée vers la décision de Stufflebeam, qui historiquement se situe avant les quatre approches décrites par Scheerens.

VI.C. L'APPROCHE "DECISIONNELLE" DE STUFFLEBEAM

Stufflebeam propose de redéfinir l'évaluation en la liant très clairement à la prise de décision. Il identifie différentes catégories de décision qui correspondent à différents types d'évaluation (voir en IV.A.2. dans le premier chapitre).

Cette proposition est bien accueillie. La présidence d'un groupe de travail financé par une fondation privée, la Phi Delta Kappa, lui permet d'approfondir et de publier sa démarche qui prendra le label CIPP (*Context, Input, Process, Product*).

Ce groupe de travail national fait d'abord le point de la question sur l'évaluation en éducation et propose ensuite une ligne de conduite. Son diagnostic des pratiques d'évaluation est sévère. L'évaluation est peu utilisée, a peu d'effets et n'influence pas l'éducation. Il propose alors aux évaluateurs d'approfondir plusieurs questions.

- Comment les évaluateurs répondent-ils aux problèmes de décision et aux demandes d'informations des différents publics, depuis les acteurs locaux jusqu'aux responsables nationaux ?
- Comment les évaluateurs reconnaissent-ils la pluralité des jugements ?

- Comment peuvent-ils adapter leur méthodologie de recherche ? La recherche et l'évaluation ont des objectifs différents qui requièrent des procédures différentes et des normes différentes. Les schémas classiques de la recherche sont inadéquats, car les problèmes de décision ont besoin de réponses immédiates et particulières et pas de connaissances généralisables.

Ce groupe de travail propose sa propre définition de l'évaluation : l'évaluation exerce une fonction par rapport à la décision, mais la décision ne fait pas partie intégrante de l'évaluation. C'est l'accent mis sur la méthodologie et la référence explicite à la décision qui permet d'intégrer ce modèle dans le type d'évaluation orientée vers la décision.

"L'évaluation en éducation est le processus par lequel on délimite, obtient et fournit des informations utiles permettant de choisir parmi les décisions possibles."

Pour construire une évaluation, l'évaluateur est face à un dilemme. D'un côté, il doit planifier soigneusement ses activités pour être efficace et crédible ; de l'autre, il lui faut rester souple et revoir périodiquement son projet pour répondre aux situations et faire une évaluation réellement répondante. C'est pourquoi il doit envisager son étude comme un processus et non comme un produit fini, il doit aussi planifier le rapport d'évaluation et les moyens de soutenir l'utilisation des conclusions.

Stufflebeam indique l'intérêt et les limites de son modèle.

- Les évaluations jouent un rôle vital dans la stimulation et la planification des changements.
- Le développement d'un programme nouveau demande une évaluation intégrée et orientée vers l'implantation.
- Mais l'évaluation est indissociable de la régulation de tout programme et n'est donc pas une activité spécialisée réservée aux projets novateurs.
- Les modèles comme le CIPP ne sont qu'une partie de la mosaïque d'évaluations formelles et informelles dans une institution.
- La mise en œuvre des évaluations ne s'indique que s'il y a un besoin d'informations évaluatives spécifiques.
- Les décisions pour commencer, soutenir, installer, arrêter un programme ou l'améliorer sont le résultat de dynamiques rationnelles et irrationnelles, qui vont au-delà de l'influence et des travaux de l'évaluateur.

VI.D. L'APPROCHE "CENTREE SUR L'UTILISATION"

Patton (1978) situe sa réflexion dans le contexte historique de l'émergence de la recherche évaluative.

Dans les années soixante, une demande d'évaluation systématique de l'efficacité des programmes gouvernementaux apparaît. C'est une réponse politique à la demande d'une plus grande responsabilité financière du gouvernement (*accountability*). Rapidement le constat est fait que l'évaluation n'influence pas les décisions de manière significative. Cette évidence provoque de nombreuses interrogations et motive les travaux de Patton et d'autres chercheurs comme Alkin ou Weiss.

Que faire pour produire de bonnes évaluations qui soient utilisées ? Telle est la question que pose Patton. Il ne suffit pas de conseils tels que *"travailler étroitement avec les décideurs"* ou *"maintenir une distance pour garantir l'objectivité et la neutralité"*. Il faut des principes pragmatiques pour concevoir et mener des évaluations spécifiques. C'est le thème du premier ouvrage de Patton.

L'originalité de la démarche de Patton tient à l'approche empirique qui la fonde. Pour développer l'utilisation, il faut d'abord la définir, puis la décrire. En 1975, Patton décide d'étudier des exemples d'utilisation. Partant de la vision pessimiste prévalant alors, il fait l'hypothèse de l'absence d'impact significatif.

L'image obtenue est un peu différente, à la fois plus complexe et moins décourageante. L'image prédominante de non-utilisation est principalement due à une définition trop étroite centrée sur l'effet immédiat, direct de l'évaluation sur les décisions du programme.

De cette étude, Patton va aussi retirer quelques principes pour mener autrement les évaluations et améliorer leur utilisation.

A la même époque et de manière indépendante, Alkin (1979) suit le même parcours empirique et arrive aux mêmes conclusions. Il propose un canevas d'intervention pour l'évaluation centrée sur l'utilisation.

Ses recommandations sont les suivantes :

- Identifier les décideurs de manière personnelle et pas comme un public abstrait.
- Identifier et préciser les questions d'évaluation pertinentes.
- Développer une attitude "active / réactive / adaptative" aussi flexible que possible face aux changements dans les buts et priorités politiques d'un programme.
- Ne pas se préoccuper de mesures quantitatives et de schémas expérimentaux, mais décrire les variations et les processus d'implantation au moyen de méthodes qualitatives ou "naturalistes".
- Tenter de construire une théorie de l'action du programme en termes de relation moyens - fins.
- Stimuler la participation des décideurs et praticiens dans les choix méthodologiques afin de les aider à comprendre et à croire dans les données ; les impliquer dans la publication des conclusions.

Dans l'approche centrée sur l'utilisation, l'évaluateur ne manipule pas le décideur et n'impose pas ses vues sur la qualité et l'utilité d'une recherche, mais il est un technicien au service du décideur.

Ce type d'approche centrée sur l'utilisation rassemble un groupe de personnes activement engagées sur le devenir du programme, son existence, les moyens de l'améliorer et de définir les besoins en informations pour réduire les incertitudes sur les effets et l'implantation.

Le pouvoir de l'évaluation est de mobiliser les énergies individuelles pour l'action. C'est le rôle du facteur personnel, qui est identifié par Patton dans sa première étude et qui est au centre de son modèle.

VI.E. L'APPROCHE "PAR LES ENJEUX"

1. L'analyse de Weiss.

Weiss (1986) présente le modèle d'évaluation "par ceux qui tiennent les enjeux".

Au milieu des années septante, de nombreuses critiques sont formulées envers les pratiques d'évaluation et leur manque de résultats. A cette époque, **trois types de réformes** sont tentés pour répondre à ces critiques dans différentes agences fédérales américaines.

- Sensibiliser les évaluateurs aux besoins locaux.

Pour augmenter l'utilisation des résultats, les évaluateurs sont priés d'impliquer les groupes d'utilisateurs dans le développement et la conduite des études d'évaluation. Ils doivent procurer des informations de façon répétée et plus volontaire.

- Financer des évaluations séparées pour différents groupes d'utilisateurs.

Plusieurs agences reconnaissent l'existence de besoins différents entre décideurs politiques fédéraux et gestionnaires de programme. Ces expériences connaissent certains succès bien que la qualité méthodologique des évaluations locales soit jugée inconstante.

- Discuter systématiquement des valeurs des groupes d'usagers et des priorités d'informations.

Des procédures pour déterminer les caractéristiques des valeurs de chaque groupe et leur poids respectif ont été mises au point. L'utilité de cette démarche a été testée, mais elle demande un temps considérable et beaucoup de participation, sans toujours satisfaire les participants. Les données recueillies ne donnent pas un résultat estimé supérieur à celles rassemblées sur une base plus intuitive ou négociée.

Le concept d'évaluation "par ceux qui tiennent les enjeux" (*Stakeholders*) a été développé par le NIE (*National Institute of Education*) sur la base de ces expériences. Le NIE est l'agence fédérale américaine qui finance les programmes novateurs en éducation et leur évaluation. L'approche par les enjeux rassemble en un tout cohérent l'implication des personnes tenant différentes positions dans la structure sociale des programmes, l'attention à leurs intérêts et besoins, une représentation de leurs priorités, un accent sur le feedback et la dissémination.

Weiss (1986) fait **la liste des hypothèses sous-jacentes à l'approche par les enjeux** et analyse celles-ci. La plupart de ces hypothèses peuvent également être appliquées à l'approche centrée sur l'utilisation proposée par Patton.

- a) Les groupes des personnes qui tiennent les enjeux peuvent être identifiés avant le début de l'évaluation.

L'identification et la représentativité des différents groupes semblent possibles dans des conditions normales. Il est cependant difficile d'envisager tous les utilisateurs potentiels, par exemple ceux qui adopteront le programme plus tard s'il réussit. La représentativité peut aussi évoluer dans le temps et dépendre de ceux qui définissent les groupes.

- b) Les personnes concernées souhaitent une évaluation à laquelle ils seraient associés.

Le souhait de participation dépend des groupes et du moment. Cependant une méfiance se manifeste généralement vis-à-vis de l'évaluation, qui apparaît plutôt comme une menace et qui n'est acceptée que si elle est la condition à un financement.

- c) Ils souhaitent des informations particulières pour les aider à planifier et à faire des choix.

Pour Weiss, les groupes d'acteurs ne sont pas vraiment des sources d'informations sur les priorités. De plus, mettre l'accent sur la flexibilité, la réponse aux besoins d'information émergents n'est pas compatible avec une évaluation sommative, orientée vers les buts, qui définit à l'avance ses indicateurs. L'évaluation qualitative est plus à même de suivre les questions qui émergent. Comment gérer des besoins contradictoires entre une information valide sur les résultats et une information sur des questions pratiques ?

- d) Les évaluateurs peuvent répondre aux demandes formulées.

Est-il possible que le rôle des évaluateurs change ? Ils ne devraient plus seulement être techniciens, mais aussi gestionnaires politiques, négociateurs, éducateurs pour partager leurs connaissances !

- e) Les participants à l'évaluation développent un sentiment de propriété envers les conclusions.

L'appropriation des résultats par les participants est favorisée par un petit nombre d'utilisateurs en accord sur les questions principales. La participation peut augmenter à condition que l'étude soit conduite de manière appropriée et avec qualité, et que les intérêts des personnes et des organisations ne soient pas menacés.

- f) Les décideurs parmi les participants (politiques, administrateurs, financeurs, directeurs de programme) utilisent les résultats comme base de décision.

Augmenter l'usage dans la prise de décision ne tient pas compte du type de décision ni de la possibilité de prendre des décisions. La prise de décision est un processus qui intègre de nombreuses informations dont l'évaluation n'est qu'une des sources. L'approche par les enjeux ne changera pas ce processus et ne fera pas de l'évaluation une information directe et exclusive pour la décision.

- g) Même si les résultats ne mènent pas à des décisions spécifiques, ils enrichissent les discussions sur le futur du programme et éclairent les acteurs.

Répondante, pertinente, bien véhiculée, l'évaluation peut produire des informations utiles sur les forces et les faiblesses. Les résultats permettent des interprétations et améliorent la compréhension du programme.

La conclusion de Weiss est lapidaire. Les bénéfices escomptés de l'approche "par ceux qui tiennent les enjeux" sont modestes en regard de l'utilisation "pour prendre des décisions".

Cette approche peut élargir le type d'information, le rendre plus pertinent. Elle permet d'être plus attentif aux situations locales, de démocratiser l'accès aux informations, d'établir un équilibre entre les intérêts des divers participants. Mais elle ne peut augmenter le caractère approprié des informations. Elle ne peut augmenter l'usage visible dans des décisions spécifiques. Elle n'augmente donc pas l'utilisation, ce qui était son objectif principal. Elle entraîne de nouvelles difficultés, car elle augmente la charge des évaluateurs en temps, en ressources et en attention. Elle peut être éclairante, mais aussi faire naître l'insatisfaction et la frustration.

La critique est encore plus vive sur d'autres aspects. Le pouvoir de décision n'est pas réparti de manière égale. Jouer la participation est une illusion. La collaboration réelle sur le plan méthodologique, entre des professionnels de l'évaluation et des acteurs sans formation, est surfaite. Enfin, des résultats non valides et non fidèles sont inutiles, même si les participants en sont heureux.

Pour Weiss, il faut mieux définir la participation. L'inclusion de nombreux groupes est une tentative pour mieux distribuer l'influence, mais la planification d'une évaluation n'est pas le meilleur lieu pour cela. Les participants souhaitent avoir quelque chose à dire, mais sur l'avenir du programme et pas sur les informations à collecter. L'évaluation ne peut être un substitut à la participation politique. Serait ce une manière de perdre les acteurs dans des détails ?

Il faut aussi combiner si possible plusieurs études. Une seule étude et une seule équipe place celle-ci en arbitre et juge de la direction à prendre pour l'étude. Il vaut mieux plusieurs études, par des équipes différentes, répondant à des enjeux différents, dont on fera la synthèse et l'arbitrage.

2. Le débat de Weiss et Patton.

La position critique de Weiss vis-à-vis de l'évaluation "par ceux qui tiennent les enjeux" peut être éclairée par un débat entre celle-ci et Patton à la fin des années quatre-vingt.

Weiss et Patton étudient tous deux l'utilisation des résultats de l'évaluation, mais adoptent une position différente. Une polémique va surgir entre eux sur le rôle de l'évaluation et sur ses perspectives de développement. Alkin (1990) tente d'éclairer ce débat et ses origines.

Alkin explique leurs divergences par des raisons circonstanciées dues à la forme du débat, par des raisons tenant au cadre de travail de ces deux auteurs et par des différences conceptuelles sur la pratique de l'évaluation.

Il fait d'abord la synthèse des éléments de consensus. Weiss et Patton sont en accord sur la plupart des facteurs qui influencent l'utilisation, sur le rôle du contexte politique, sur l'influence de la connaissance de terrain du décideur / utilisateur préalable à l'évaluation, sur l'influence partielle de l'évaluation dans la décision, sur l'impact réel de l'évaluation dans le développement des programmes.

Les désaccords sont fonction d'un contexte d'application différent de l'évaluation. Pour Weiss, l'évaluation, même orientée vers l'utilisation, influence peu les décisions. Mais elle travaille à un niveau politique, fédéral, sur un petit nombre de grands programmes dirigés par des agences ou par le Congrès. Ces évaluations ont un haut niveau de visibilité, de multiples enjeux, de nombreux partisans ou détracteurs. Les décisions sont plutôt de type "cumulatif" comme le montrent ses travaux (Weiss, 1990). Quand Weiss affirme que les prescriptions de l'évaluation centrée sur l'utilisation ont peu de succès, Alkin doute qu'elle ait pu appliquer réellement ces préceptes dans les situations où elle travaille.

Patton, par contre, travaille sur des programmes beaucoup plus ciblés sur le plan local. Dans ce contexte, il paraît possible d'identifier plus clairement les "usages définis par les utilisateurs identifiés" ("*intended use by intended users*" selon l'expression de Patton). L'évaluation est alors clairement une base pour la décision et les conseils pour augmenter l'utilisation peuvent s'avérer efficaces. Les petits programmes sont moins marqués politiquement, sont moins visibles, ont moins de personnes impliquées. Les décisions portent plus sur des modifications et sont moins menaçantes. Cette évaluation plus spécifique peut entraîner plus directement des effets.

Cependant, Alkin affirme que les différences entre Patton et Weiss sont plus fondamentales. Elles sont conceptuelles et méthodologiques. Weiss est un évaluateur orienté vers la recherche "académique". Patton est un évaluateur orienté vers le service à des clients. Ils opèrent avec des paradigmes différents, mais il est difficile d'établir si ce sont les contextes qui influencent le choix de tel ou tel paradigme ou l'inverse.

A l'évidence ces paradigmes pèsent sur la façon d'aborder l'évaluation, par exemple l'identification des décideurs, mais aussi sur la manière de récolter, de traiter l'information et d'en rendre compte.

Ce débat met en lumière l'importance du contexte de l'évaluation, qui était déjà invoqué par Stufflebeam pour expliquer ses divergences avec Scriven. Le type de programmes et de décisions jouent un rôle sur les pratiques d'évaluation.

VI.F. LES MODALITES DE L'EVALUATION "PROCESSUS - DECISION"

L'examen des différentes modalités d'évaluation selon les axes interne - externe et informel - formel permet de faire la synthèse des différentes formes possibles d'évaluation "processus - décision".

Les 16 modalités de l'évaluation "processus - décision"

Jugement implicite	DECISION INTERNE				DECISION EXTERNE			
	Processus interne	Processus externe			Processus interne	Processus externe		
Processus informel	Pii	Pei	J		Pii	Pei	J	
Processus formel	Pif	Pef			Pif	Pef		
DECISION INFORMELLE	Dii				Dei			
Processus informel	Pii	Pei	J		Pii	Pei	J	
Processus formel	Pif	Pef			Pif	Pef		
DECISION FORMELLE	Dif				Def			

1. La première forme est interne tant pour le processus que pour la décision.

Le processus informel ou formel entraîne une décision immédiate ou différée. Cette situation se conçoit au sein d'un programme d'action et reste souvent implicite. L'énonciation se fera pour répondre à un besoin de communication interne ou externe : réunion d'équipe, rapport au commanditaire.

Par exemple, la décision de modifier, en cours de programme, une stratégie d'information du public sous forme de conférences (décision formelle) est justifiée dans un rapport par le constat (non formalisé) d'une faible fréquentation de ces exposés (Pii J Dif).

2. La deuxième forme contient un processus interne et une décision externe.

Par exemple, le rapport d'une évaluation interne formalisée réalisée par les acteurs d'un programme de prévention indique de manière précise plusieurs décisions à prendre par le Gouvernement afin d'améliorer l'efficacité de ce type de programme dans le futur (Pif J Dei).

3. La troisième forme prévoit un processus externe et une décision interne.

C'est de manière assez claire la situation d'une commande d'évaluation par les acteurs d'un programme en vue de résoudre un problème, de prendre une décision. L'évaluateur externe est chargé de proposer des alternatives sans nécessairement devoir passer par la démonstration de la valeur de ses conclusions.

4. La quatrième forme est entièrement externe.

Le processus, plus ou moins formalisé, et la décision, plus ou moins directe, sont entre les mains d'interlocuteurs externes au programme à évaluer.

VII. L'EVALUATION TOTALEMENT EXPLICITE

T7 : L'évaluation explicite (processus, jugement et décision explicites)

Processus	Jugement de valeur	Décision	<i>Explicite</i>
			<i>Implicite</i>

VII.A. EVOLUTION DE LA DEFINITION DE L'EVALUATION DE STUFFLEBEAM

Comme tous les modèles, celui de Stufflebeam a évolué. Il propose, en 1985, une définition plus complète de l'évaluation dans laquelle il intègre les dimensions de description et de jugement ainsi que les fonctions de justification et de compréhension. Selon lui, le modèle CIPP doit aussi pouvoir assurer une fonction sommative. L'évaluation doit non seulement servir la conduite et la résolution de problème, mais aussi fournir une base aux décisions institutionnelles, être disponible pour un examen public, comme le suggère Scriven.

Stufflebeam entend ainsi répondre à un certain nombre de critiques lui reprochant de négliger le rôle fondamental de l'évaluation par rapport au jugement. Il explique sa première approche par des raisons de contexte. Scriven a surtout travaillé au niveau fédéral pour comparer des curriculums. Stufflebeam travaillait au niveau d'un état avec des écoles.

"L'évaluation est le processus qui délimite, obtient et procure des informations descriptives et des jugements sur la valeur et le mérite de différents objets comme les buts, le plan, l'implantation et les effets, de manière à guider la prise de décision, servir les besoins de justification financière et promouvoir la compréhension des phénomènes impliqués."

Cette définition est la plus complète que l'on puisse admettre en se plaçant du point de vue d'un évaluateur externe. Elle guide la prise de décision, mais ne l'exécute pas. Elle marque la limite d'une vision professionnelle de l'évaluation. Nous la considérons cependant comme une forme de l'évaluation totalement explicite parce qu'elle se construit en référence à la décision, même si celle-ci dépend d'une autre personne.

VII.B. MONNIER

Monnier (1987) dont nous avons présenté la critique des modèles administratifs français et du modèle de l'expérimentation sociale nord-américaine, fait aussi une proposition sur la manière de gérer une évaluation de l'action des pouvoirs publics. Elle est très proche de celle proposée par Patton.

Pour Monnier (1987, p. 123), le chargé d'évaluation ne peut plus être un évaluateur traditionnel, expert en méthodes et examinateur de la valeur d'un programme. La sociologie des organisations a fait progresser la compréhension de la gestion des problèmes sociaux. Celle-ci est un processus pluraliste jamais achevé fait de négociations, de compromis entre protagonistes parfois peu complémentaires ou mêmes franchement opposés. Ce processus progresse par phases "tourbillonnaires" comprenant des ébauches de solution, des actions qui sont reformulées en cours de route et qui demandent une évolution des objectifs et des moyens, et enfin des effets qui entraînent des modifications le plus souvent implicites (Monnier, 1987, pp. 85 à 94). Dans ce cadre, le chargé d'évaluation pluraliste devient un maïeuticien, un médiateur et un méthodologue (Monnier, 1987, pp. 123 à 124).

Il est maïeuticien, car il doit clarifier les questions des acteurs. Il doit dévoiler les enjeux sociaux et les valeurs, préciser les critères utilisables. Il doit engager les acteurs dans une démarche d'analyse et de compréhension de la situation préalable à l'évaluation. L'évaluateur agit déjà en permettant une modification des représentations. Cette procédure a déjà été présentée dans les modèles d'évaluation de Stake, Cronbach, Patton.

Il est méthodologue parce que son intervention vis-à-vis des acteurs doit être accompagnée, supportée par des informations les plus valides possibles ; l'évaluateur est souvent un garant de la crédibilité externe d'une action aux yeux de l'autorité publique. C'est aussi une fonction bien documentée par les théoriciens de l'évaluation.

Il est enfin médiateur et c'est l'apport le plus original de Monnier. L'évaluateur est engagé dans le processus politique, comme l'affirmait Cronbach, mais pas de manière "neutre" ; il s'engage comme médiateur pour obtenir des compromis sur la définition des objectifs de l'évaluation, sur les informations prioritaires à recueillir. Il doit aussi découvrir et formuler les conclusions, les propositions qui permettent une approbation large et qui sont des compromis aussi pour la poursuite de l'action. L'évaluateur, selon notre grille d'analyse, devient un co-décideur.

La proposition de Monnier montre donc un évaluateur au coeur de l'action, comme l'est l'évaluation.

VII.C. VIAL

Cette évolution permet de mieux saisir l'unité interne du processus d'action, de la prise d'information jusqu'à la décision en passant par le choix entre des alternatives d'action. La proposition de Vial (1987) d'une évaluation "formatrice" portant sur les processus d'apprentissage va plus loin dans ce sens. Elle rencontre notre proposition d'une nature continue de l'acte évaluatif au sein de l'action.

Pour Vial, l'enjeu de l'évaluation n'est pas le contrôle ou le soutien de l'extérieur, mais sa maîtrise par l'apprenant. L'évaluation "formatrice" vise à fournir à celui-ci les moyens de d'explicitier et de formaliser les critères qui lui sont nécessaires pour évaluer son apprentissage.

L'apport des vues piagésiennes, notamment sur le statut de l'erreur et l'importance de l'apprentissage actif n'a pas été transposé à l'évaluation de programmes. Le fait d'assimiler le gestionnaire d'un programme à un apprenant, de situer l'évaluation par rapport à l'action dans sa dimension interne, peut élargir et enrichir la notion d'auto-évaluation qui n'a été qu'effleurée et plutôt dévalorisée par la plupart des évaluateurs professionnels.

VII.D. LES MODALITES DE L'EVALUATION EXPLICITE

Le grand nombre de modalités théoriques dans un système à trois composantes, qui peuvent chacune prendre quatre positions différentes, ne doit pas rebuter. Il reflète la diversité des situations possibles. La description et l'analyse des pratiques rencontrées permettra d'estimer si certaines combinaisons sont purement théoriques.

Les 64 modalités d'évaluation explicite

		INTERNE				EXTERNE				
		INTERNE		EXTERNE		INTERNE		EXTERNE		
		interne	externe	interne	externe	interne	externe	interne	externe	
INF	INF	inf	Pii	Pii	Pii	Pii	Pei	Pei	Pei	Pei
			Jii	Jii	Jej	Jej	Jii	Jii	Jej	Jej
			Dii	Dei	Dii	Dei	Dii	Dei	Dii	Dei
INF	INF	for	Pii	Pii	Pii	Pii	Pei	Pei	Pei	Pei
			Jii	Jii	Jej	Jej	Jii	Jii	Jej	Jej
			Dif	Def	Dif	Def	Dif	Def	Dif	Def
INF	FOR	inf	Pii	Pii	Pii	Pii	Pei	Pei	Pei	Pei
			Jif	Jif	Jef	Jef	Jif	Jif	Jef	Jef
			Dii	Dei	Dii	Dei	Dii	Dei	Dii	Dei
INF	FOR	for	Pii	Pif	Pif	Pif	Pei	Pei	Pei	Pei
			Jif	Jif	Jef	Jef	Jif	Jif	Jef	Jef
			Dif	Def	Dif	Def	Dif	Def	Dif	Def
FOR	INF	inf	Pif	Pif	Pif	Pif	Pef	Pef	Pef	Pef
			Jii	Jii	Jej	Jej	Jii	Jii	Jej	Jej
			Dii	Dei	Dii	Dei	Dii	Dei	Dii	Dei
FOR	INF	for	Pif	Pif	Pif	Pif	Pef	Pef	Pef	Pef
			Jii	Jii	Jej	Jej	Jii	Jii	Jej	Jej
			Dif	Def	Dif	Def	Dif	Def	Dif	Def
FOR	FOR	inf	Pif	Pif	Pif	Pif	Pef	Pef	Pef	Pef
			Jif	Jif	Jef	Jef	Jif	Jif	Jef	Jef
			Dii	Dei	Dii	Dei	Dii	Dei	Dii	Dei
FOR	FOR	for	Pif	Pif	Pif	Pif	Pef	Pef	Pef	Pef
			Jif	Jif	Jef	Jef	Jif	Jif	Jef	Jef
			Dif	Def	Dif	Def	Dif	Def	Dif	Def

Nous ne reprendrons pas en détail les différentes modalités possibles de l'évaluation explicite. L'évaluation totalement implicite ou totalement explicite sont les deux extrêmes d'un continuum qui permet de mettre en oeuvre trois indicateurs des pratiques d'évaluation : l'énonciation, la position interne ou externe des acteurs et la formalisation des trois composantes de l'acte évaluatif (le processus, le jugement et la décision).

Toutes les formes d'évaluation, les modèles que nous avons découverts au fil de ce chapitre représentent des positions intermédiaires sur ce continuum. Une évaluation implicite de gestion, une expertise, une évaluation "scientifique", une évaluation d'accréditation, une évaluation commanditée sur la valeur d'un programme, une évaluation en interaction avec le décideur, une évaluation "médiatrice" ou "formatrice", toutes peuvent à un moment ou à un autre devenir totalement implicites ou explicites selon les circonstances qui nécessitent ou non l'énonciation.

Souvent, la difficulté de situer très précisément un modèle est le signe que l'approche uniquement théorique, académique est insuffisante. Entre le discours de principe et les réalités des enjeux, les situations concrètes de terrain, se déploie une distance qu'il reste à explorer.

Reconnaître la diversité, la comprendre au moyen d'une grille descriptive, en identifier les conditions pour, plus tard, pouvoir l'exploiter au mieux et répondre aux besoins d'évaluation des acteurs de l'éducation pour la santé. Tel est notre but.

CHAPITRE 3

LES RECHERCHES SUR L'UTILISATION ET LA PRATIQUE DE L'EVALUATION

I.	L'UTILISATION DE L'EVALUATION110
I.A.	PATTON (1978) : L'UTILISATION, UNE PERSPECTIVE EMPIRIQUE111
I.B.	ALKIN, DAILLAK, WITHE (1979) : L'EVALUATION FAIT-ELLE UNE DIFFERENCE ?113
I.C.	UNE REVUE DE LITTERATURE120
II.	LA PRATIQUE DE L'EVALUATION132
II.A.	L'ETUDE DES THEORIES132
II.B.	L'ETUDE DES PRATIQUES D'EVALUATION137

Les années soixante ont vu l'émergence de la perception de l'inefficacité de l'évaluation et du lien étroit que celle-ci doit établir avec l'action et la prise de décision. A partir de la fin des années septante, les études sur l'utilisation de l'évaluation entament leur développement, avec les travaux de Patton et d'Alkin, qui fournissent ainsi une base empirique à leurs modèles.

Les résultats de ces recherches sont présentés dans une série de revues de littérature que nous parcourons dans la première partie de ce chapitre. Ils ont permis de faire une série de recommandations pour augmenter le potentiel d'utilisation des évaluations.

A la fin des années quatre-vingt, Alkin et Coyle (1988) font le point de la réflexion sur l'utilisation, la non utilisation et les mauvaises utilisations de l'évaluation. Alkin et Coyle souhaitent élargir le débat à l'utilisation appropriée ou non de l'évaluation et plus seulement à son degré d'utilisation, c'est-à-dire qu'ils veulent lancer une réflexion sur la bonne ou la mauvaise utilisation en relation avec la bonne ou la mauvaise pratique de l'évaluation.

Les études américaines entreprises sur l'évaluation se préoccupent de l'efficacité de l'évaluation, de l'utilisation des résultats de l'évaluation. Elles s'intéressent fort peu au processus global de l'acte évaluatif, si ce n'est sous l'angle de la méthodologie. On trouve peu de recherches sur les pratiques et, quand il y en a, elles portent sur les pratiques des professionnels de l'évaluation. Nous les aborderons dans la deuxième partie de ce chapitre.

Il peut paraître étonnant que les chercheurs qui ont étudié l'utilisation de l'évaluation ont considéré comme acquise l'existence même de l'évaluation et n'aient pas posé la question préalable des conditions d'apparition de la pratique de l'évaluation. Les conditions socioculturelles, économiques et politiques entourant l'existence de l'évaluation aux Etats-Unis en sont probablement responsables. L'évaluation n'est pas chez nous une pratique socio-politique aussi évidente.

Les études sur les facteurs explicatifs de l'utilisation des résultats des évaluations se sont donc surtout préoccupées des relations entre les évaluations externes, réalisées par des professionnels, et les décisions administratives et politiques. Notre projet est plus global. Sur la base de l'unité constitutive de l'évaluation, nous analysons les pratiques au sein des programmes, qu'elles soient internes ou externes, formalisées ou non, pourvu qu'elles soient énoncées socialement.

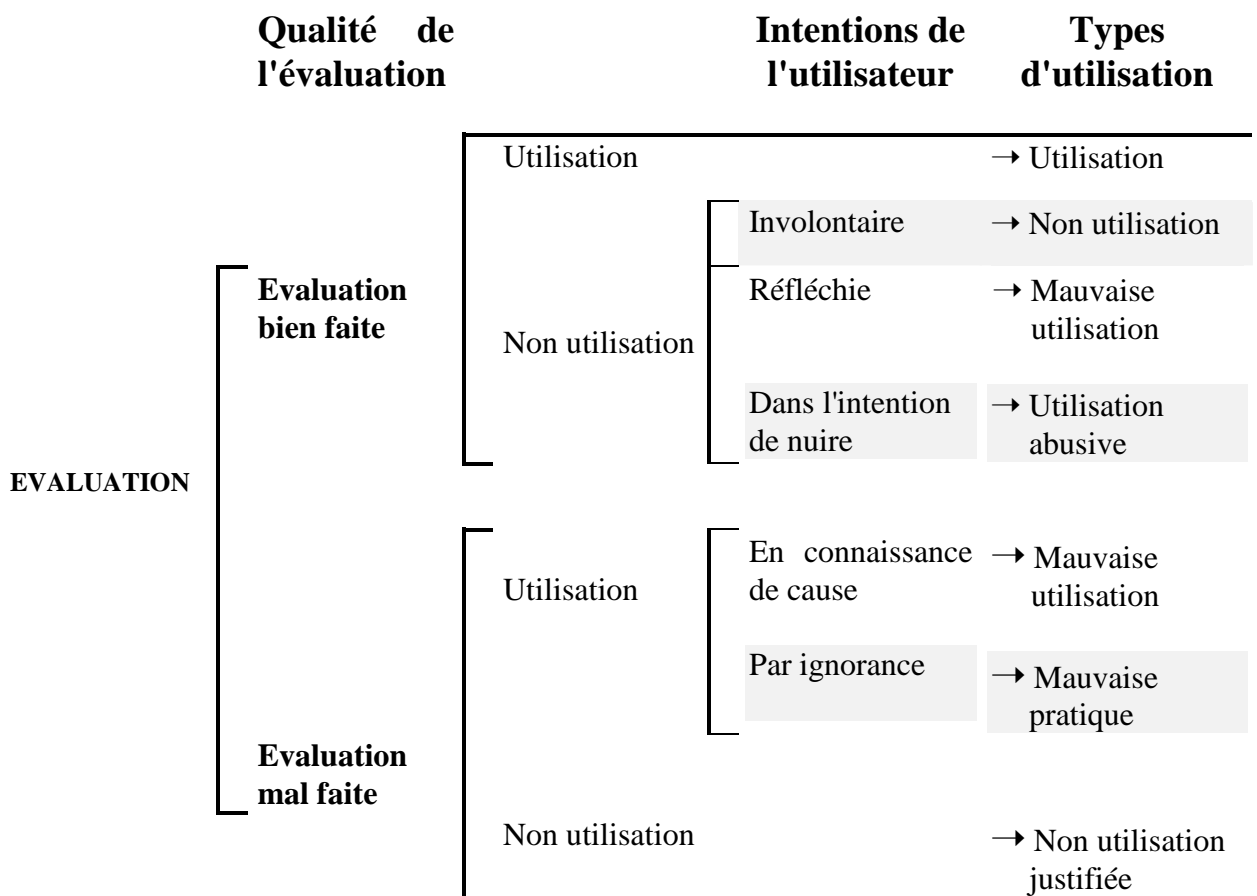
Cependant, il est intéressant de prendre connaissance de ces études sur l'utilisation pour les raisons suivantes :

- ce sont les seules recherches sur l'évaluation qui aient été développées suffisamment pour conduire à des revues de littérature ;
- ces études ont tenté de clarifier le concept d'utilisation, qui est une des composantes de la définition de l'évaluation que nous adoptons ;
- notre méthodologie s'apparente aux méthodes qualitatives et rétrospectives utilisées par ces études ;
- enfin, les résultats de celles-ci mettent en évidence le rôle de facteurs antérieurs à l'utilisation, lors de la conception et de la mise en œuvre de l'évaluation.

I. L'UTILISATION DE L'EVALUATION

Dans cette première partie, nous aborderons les concepts, les méthodes de recherches et les résultats de celles-ci en nous appuyant sur des revues de littérature (Alkin, 1985, Alkin et Daillak, 1990, Cousins et Leithwood, 1985, Leviton et Hughes, 1981).

Pour mettre en perspective ces débats, Alkin et Coyle (1988) proposent un schéma des relations possibles entre plusieurs types d'utilisation, en fonction de deux critères : la qualité de la réalisation de l'évaluation et l'intention de l'utilisateur.



Alkin et Coyle font d'abord la distinction entre la pratique de l'évaluation qui est sous la responsabilité de l'évaluateur et l'utilisation qui est de la responsabilité du décideur.

La responsabilité de l'évaluateur est une responsabilité méthodologique. Les mauvaises pratiques sont :

- le manque de prise en compte des besoins en information des utilisateurs,
- des fautes dans le recueil et l'interprétation des données,
- des erreurs dans la communication des conclusions.

Il y aura donc des évaluations bien faites et d'autres mal faites.

La responsabilité de l'utilisateur va varier en fonction de la qualité de l'évaluation, de sa capacité à comprendre les erreurs méthodologiques et de ses intentions d'utilisation.

- Si l'évaluation est mauvaise, la non utilisation est justifiée ou l'utilisation par ignorance est de la responsabilité de l'évaluateur. Si le décideur utilise une mauvaise évaluation en connaissance de cause, il s'agit d'une mauvaise utilisation.
- Si l'évaluation est bonne, la non utilisation involontaire est simplement une absence d'utilisation. Par contre, la non utilisation volontaire est un abus plus ou moins important selon la gravité des intentions.

Alkin et Coyle donnent ensuite quelques exemples de mauvaise utilisation de l'évaluation. Il peut y avoir des abus :

- dans la demande d'évaluation : *pour des avantages politiques, financiers, relationnels ;*
- dans le déroulement de l'évaluation : *retarder le programme, éviter des responsabilités ;*
- dans l'utilisation des conclusions : *les changer, les tronquer, les ignorer, les simplifier.*

La proposition de Alkin et Coyle est bien sûr critiquable parce qu'elle repose sur une simplification des rapports entre acteurs. L'évaluateur a lui aussi des intérêts à réaliser une évaluation de bonne ou de mauvaise qualité, et pas seulement sous la pression des décideurs. Cette proposition est surtout intéressante par l'image qu'elle donne d'une "bonne" pratique méthodologique d'évaluation.

Avant de commenter les revues de littérature sur l'utilisation, nous proposons une vision plus approfondie des deux premières recherches sur le thème de l'utilisation, réalisées par Patton et ses collaborateurs (1978) et Alkin, Daillak et White (1979). Outre leur intérêt historique, leur caractère exemplaire par rapport aux travaux de ce type et leur spécificité dans le domaine de l'évaluation, elles nous apportent des éléments méthodologiques précieux, car elles ressortissent des mêmes options que notre étude : qualitative et rétrospective.

I.A. PATTON (1978) : L'UTILISATION, UNE PERSPECTIVE EMPIRIQUE

Pour Patton, la question principale n'est pas la naissance d'une évaluation, mais son devenir, ce qui se passe ensuite. L'existence de l'évaluation fait-elle une différence ? Y a-t-il un impact sur l'environnement ?

En 1975, Patton décide d'étudier des exemples positifs d'utilisation en collaboration avec quelques étudiants d'une formation en méthodologie de l'évaluation. Ils souhaitent mesurer le degré d'utilisation et identifier les facteurs qui affectent ces variations. Ils partent de la définition par les usagers eux-mêmes de ce qu'est l'utilisation. Ces usagers sont des décideurs au niveau fédéral dans le secteur de la santé mentale.

1. Méthode.

L'échantillon est constitué de 20 cas extraits d'un ensemble de 170 études financées par l'agence fédérale "*Office of Health Evaluation*" entre 1971 et 1973. Ces 20 cas concernent des centres communautaires de santé mentale, de lutte contre l'alcoolisme, des actions de quartier, des programmes de formation, ... Ils vont d'une évaluation interne d'une durée de trois semaines à un programme d'évaluation de quatre ans. Quatorze sont externes et 6 internes.

Les entretiens sont intensifs auprès de trois personnes clés par cas : le responsable du projet, la personne identifiée par ce responsable comme étant le décideur ou la personne connaissant le mieux l'impact de l'étude et l'évaluateur qui a la responsabilité principale.

La question est : "Quel a été l'impact de l'étude d'évaluation sur le programme de votre point de vue : l'impact sur les opérations du programme, sur la planification, sur le financement, sur la politique, sur les décisions, sur les représentations du programme ?"

Selon les réponses, une autre question était posée ensuite : "Certaines études ont un impact plus large qu'immédiat comme la façon de penser les problèmes, la législation, ... Y avait-il un impact de cette sorte ?"

Douze personnes réalisent les entretiens et partagent leurs perceptions pour en établir l'analyse. Les exemples de confirmation et d'infirmité des hypothèses sur les thèmes dominants seront réexaminés de manière indépendante par d'autres membres pour aboutir à un consensus.

2. Résultats.

Septante-huit pour cent des décideurs et 90 % des évaluateurs trouvent qu'il y a eu un impact sur le programme. Quatre-vingts pour cent des décideurs et 70 % des évaluateurs peuvent identifier aussi l'absence d'impact.

Selon les auteurs, ce nombre considérable et étonnant de réponses positives par rapport aux résultats des autres études de l'époque s'explique par la manière de définir l'impact. L'image prédominante de non utilisation est due à une définition trop étroite centrée sur l'effet immédiat, direct, concret sur les décisions de programme.

L'évaluation ne mène pas directement et immédiatement à des décisions concrètes et décisives, sauf exception. L'impact trouvé correspond à une information supplémentaire qui est jugée et utilisée dans le contexte d'autres informations disponibles pour aider à réduire les inconnues dans le processus d'une décision difficile.

Le degré d'influence est donc difficile à déterminer surtout quand l'étude a été longue. Les changements, même s'ils vont dans le sens des recommandations de l'évaluation, ne sont pas simplement et directement imputables à celle-ci. Les difficultés sont perçues progressivement et avant l'évaluation. Celle-ci les confirme. Elle apporte une crédibilité supplémentaire. Le décideur est plus confiant et déterminé, un changement peut se mettre en place plus rapidement. Ce sont des impacts jugés importants même s'il ne sont pas décisifs, s'ils ne correspondent pas à des impacts majeurs sur des décisions concrètes.

L'évaluation ne permet pas de rationaliser rapidement un système. L'évaluation agit comme la recherche qui suit le plus souvent une voie lente, graduelle.

3. Conclusions.

La méthodologie, rétrospective et qualitative, semble bien permettre une analyse assez riche de l'impact d'un programme auprès des personnes les plus directement impliquées.

Cette première étude permet d'emblée d'élargir la définition de l'utilisation. Patton met en évidence une influence de type cognitif qui s'insère graduellement dans un contexte décisionnel.

I.B. ALKIN, DAILLAK, WITHE (1979) : L'EVALUATION FAIT-ELLE UNE DIFFERENCE ?

Basé sur une analyse systématique et approfondie de six cas, le travail de Alkin, Daillak et White tente de répondre à la question de l'utilisation de l'évaluation.

1. Définition : qu'est ce que l'utilisation ?

Comment reconnaître une utilisation, comment en définir la fréquence ? Pour Alkin et ses collègues, la réponse théorique à cette question dépend de la définition donnée au concept d'"utilisation".

Deux positions simples définissent les deux extrémités d'un continuum. D'une part, l'utilisation est liée à une décision spécifique, immédiate d'un utilisateur unique et intentionnel. Elle est entièrement fondée sur le rapport d'évaluation. D'autre part, l'utilisation peut être le fait de n'importe qui à n'importe quel moment pour n'importe quel objectif.

Pour sortir de cette impasse, il est nécessaire d'isoler les composants essentiels de l'utilisation : la communication d'une information évaluative, le caractère approprié de l'utilisateur, le contexte de l'information (isolée ou non) et enfin l'utilisation proprement dite pour des actions, des décisions ou des modifications d'attitudes.

De l'analyse de ces éléments émerge une définition complexe de l'utilisation.

"Une forme d'utilisation de l'évaluation d'un programme scolaire local, c'est une information considérée par un client local, des usagers locaux autorisés ou des usagers externes influents comme un élément d'influence dominant, partagé ou cumulatif dans la prise de décisions, la confirmation de décisions ou d'actions préalables, l'établissement ou la modification d'attitudes en relation avec l'établissement, le prolongement d'un financement externe ou local, le maintien d'une composante des méthodes d'instructions, des opérations vis-à-vis du personnel ou l'acceptation de la communauté vis-à-vis d'un programme scolaire local." (Alkin, Daillak et White, 1979)

2. Méthode.

L'émergence du schéma d'interprétation a suivi un long cheminement fait d'aller et retour entre les données et leur analyse soutenue par l'expérience des auteurs et la vision de la situation des informateurs.

Après une phase de recueil des faits et d'interprétation, chaque membre de l'équipe de recherche a développé, pour chaque cas, une "carte conceptuelle" de chaque facteur d'utilisation en tentant de rassembler les éléments clés comme les événements critiques, les facteurs contextuels, etc. Ces cartes ont ensuite été rassemblées cas par cas pour obtenir un pattern clarifiant le réseau des éléments.

A l'étape suivante, les différents schémas des différents cas ont été confrontés pour en déterminer les similarités et produire un schéma commun qui est à la base d'une catégorisation de facteurs.

3. Résultats.

Alkin, Daillak et White remarquent que le degré d'utilisation de l'évaluation n'est pas le fait du hasard, mais qu'il est associé à certaines caractéristiques de la situation d'évaluation. Ils proposent une classification des facteurs d'utilisation, qui comporte huit catégories et de nombreuses sous-catégories.

Classification des facteurs d'utilisation

- a) Le contexte préexistant à l'évaluation
- b) Les orientations des utilisateurs
- c) Les approches de l'évaluateur
- d) La crédibilité de l'évaluateur
- e) Les facteurs organisationnels et institutionnels internes
- f) Les facteurs institutionnels externes
- g) Le contenu de l'information et le rapport
- h) Le style de l'administrateur

a) Le contexte préexistant à l'évaluation.

L'évaluation est un processus dynamique dans lequel les activités et les perceptions des participants jouent un rôle important influençant l'évaluation elle-même et son utilité. Avant même sa mise en place effective, des événements vont influencer l'ampleur et la direction de l'évaluation.

Tout d'abord intervient le contexte, comme les caractéristiques de la communauté ou les problèmes qui s'y posent. Ensuite viennent les contraintes fixées à l'évaluation par des contrats écrits, comme l'exigence d'une évaluation du rendement de l'élève par des tests, ou d'autres éléments venant de différents niveaux de décision (fédéral, local, ...). Enfin viennent les contraintes financières, rarement sensibles pour les gestionnaires de programme, mais qui fixent une limite tangible pour l'évaluateur. Ces éléments de départ - contexte, contrat, et budget - vont influencer les choix d'évaluation.

b) Les orientations des utilisateurs.

Les intérêts et les attitudes des utilisateurs occupent une position centrale dans une théorie de l'utilisation.

Les questions des différents utilisateurs sont les points de départ de l'évaluation. Elles influenceront l'utilisation postérieure des informations. Elles peuvent être plus ou moins claires ou pressantes. L'évaluateur peut ou non en favoriser l'expression. Elles sont très variables selon les catégories d'utilisateurs.

Les usagers ont aussi une idée plus ou moins précise de ce qu'est l'évaluation et de ce qu'elle peut apporter. *Dans un cas, par exemple, le directeur veut clairement des données "objectives" pour apprécier les résultats et l'évaluateur se centre sur les effets. Dans un autre, l'évaluation est vue comme un mal nécessaire et même susceptible d'influer négativement sur la coopération.*

L'attitude de l'utilisateur vis-à-vis de la forme prise par l'information joue aussi un rôle. *Par exemple, les données quantitatives peuvent entraîner différentes réactions : un malaise, un détournement vers d'autres informations ou à l'inverse, une augmentation du prestige et de la crédibilité de l'évaluateur.* Certains préfèrent des données brutes, d'autres des recommandations directement utilisables.

c) Les approches de l'évaluateur.

Les six cas suggèrent que la façon dont l'évaluateur conçoit et définit son travail influence l'utilisation. L'information produite par les évaluateurs est utilisée avec plus ou moins de succès selon les circonstances et certaines caractéristiques de leur approche de l'évaluation.

Une bonne partie de la littérature sur l'évaluation est consacrée à des modèles. Aucun des cas analysés n'applique de tels modèles et cette observation est assez généralisable. Non pas que les évaluateurs les ignorent, mais ils ont développé leur style et leurs méthodes.

Un autre leitmotiv de la littérature est l'attention aux considérations méthodologiques de recherche et d'analyse, au besoin de rigueur. Les six cas sont assez exemplaires des difficultés dénoncées dans la littérature. Un seul utilise une démarche quasi-expérimentale et les autres correspondent aux devis préexpérimentaux proposés par Campbell et Stanley (1963). La rigueur relative de l'une des évaluations a été un élément favorable dans l'utilisation, mais il n'y a pas d'évidence que le manque de rigueur des autres évaluations ait joué contre leur utilisation.

En résumé, la rigueur semble moins importante pour l'utilisation que ce que suggère la littérature. Du point de vue des utilisateurs toutes les procédures paraissent rigoureuses et c'est leur point de vue qui fait la différence en utilisation. L'approche rigoureuse d'un des cas correspond à la demande du décideur. Dans les autres cas, il n'y a pas une telle demande.

Le choix d'un rôle par l'évaluateur est toujours possible dans le cadre qui lui est imparti. Il sera fonction de son expérience, de son style, de sa formation, combinant des considérations professionnelles et personnelles. En fonction de ce rôle et d'autres attitudes envers l'évaluation, les évaluateurs peuvent avoir des positions divergentes sur la désirabilité de l'engagement des usagers : une participation active ou une présence limitée et contrôlée dans le processus. De leur côté, les utilisateurs potentiels peuvent être plus ou moins intéressés par une telle participation.

Les cas présentent aussi des différences dans la manière dont l'évaluateur conçoit son rôle dans la diffusion de l'information. Les données peuvent être présentées et expliquées à différents moments d'une évaluation. L'évaluation ne se termine pas quand le rapport est remis. La stimulation et la facilitation consistent à discuter les conclusions, à aider à tirer les implications et les recommandations pour l'action, à contrôler les résultats des modifications sur la base de l'évaluation. La création de rapports étroits, personnels et professionnels, renforcent la possibilité de suggérer des utilisations de l'évaluation.

d) La crédibilité de l'évaluateur.

L'impact de l'évaluation dépend de la façon dont l'utilisateur fait confiance à l'évaluateur. Les procédures sont-elles correctes ? Puis-je avoir confiance si ces techniques d'évaluation dépassent mes compétences ? Le jugement personnel de l'utilisateur jouera un rôle important. La confiance et la crédibilité sont des matières fluctuantes en fonction des sujets traités, des moments et des utilisateurs. Elles sont liées à la perception des compétences techniques sur la collecte et l'analyse de données, mais aussi des compétences sur le sujet ou de l'expérience pratique. La crédibilité peut aussi évoluer et s'étendre progressivement à d'autres sujets ou publics.

L'autonomie du programme est plus ou moins grande ; la position du directeur du programme peut varier ainsi que celle de l'évaluateur. L'évaluateur désigné ou choisi doit commencer par gagner la confiance et convaincre du rôle de l'évaluation.

e) Les facteurs organisationnels et institutionnels internes.

Cette rubrique concerne les facteurs qui sont liés directement au programme sur le site local.

Entre l'organisation d'une école et d'un programme et la mission définie à l'évaluation existe un espace de liberté pour mener des évaluations spécifiques au site qui auront une probabilité d'utilisation supérieure. Chaque facteur qui influence la conduite d'une évaluation peut influencer l'utilisation de l'information. Dans les cas étudiés, les enseignants et l'équipe du programme influencent parfois quoique rarement la conduite de l'évaluation.

La convergence entre les informations de l'évaluation et les autres sources d'information joue un rôle considérable dans l'acceptation et l'utilisation.

Il y a des risques personnels et institutionnels ainsi que des bénéfices potentiels à mener une évaluation. Ces coûts et ces bénéfices influencent les évaluateurs et les utilisateurs et, par conséquent, la production et l'utilisation des informations. Dans certains cas, l'évaluation est facilitée parce que les risques sont minimales sur le plan personnel et institutionnel (position stable) et que les avantages sont grands d'une valorisation supplémentaire.

f) Les facteurs institutionnels externes.

La nature et l'évolution des relations entre la communauté, le programme, l'école, le district, l'état fédéral, entre les différentes agences vont influencer la conduite de l'évaluation et l'usage qui en sera fait.

Dans la plupart des cas, la communauté exerce une influence au départ de l'évaluation, puis reste assez neutre. Dans un des cas, elle joue un rôle dominant et actif. *L'éducation et les résultats des enfants sont un véritable enjeu d'intégration pour une communauté minoritaire et ses craintes vis-à-vis de l'évaluation sont élevées. La pression politique est importante. Une fois la communauté rassurée par des résultats positifs, la pression diminue.*

Le rôle de l'évaluation dans le renouvellement du financement du programme influence sa conduite et sa perception. L'influence peut être statique (les exigences extérieures sont stables, l'évaluation ne les dépasse pas) ou dynamique (les exigences changent en cours de programme et l'évaluation en tient compte ou parfois, elle les anticipe).

g) Le contenu de l'information et le rapport.

L'information évaluative ne se réduit pas à un rapport final écrit contenant des données quantitatives, des analyses statistiques, un résumé des conclusions sur la valeur, l'efficacité et la relative opportunité de certaines décisions. L'information peut varier dans sa forme, son contenu, sa fréquence.

L'évaluateur doit faire appel à ses compétences personnelles et professionnelles pour déterminer les destinataires de l'information et sélectionner le contenu et la forme de cette information parmi toutes celles disponibles. Une information bien ciblée sur les questions des utilisateurs aura plus de chance d'être utilisée. La forme importe autant que le contenu. Les préférences personnelles sont importantes. Par exemple, une information correcte et adéquate, mais trop quantitative conduit un directeur à s'appuyer plus sur ses propres observations et conversations.

Mais ces dimensions ne sont pas intéressantes en elles-mêmes pour expliquer l'utilisation. Elles doivent être situées dans le contexte particulier de l'évaluation, qui intègre les utilisateurs et leurs perceptions. C'est le processus d'échange d'informations qui importe. Ce qu'il faut, ce n'est pas un simple échange d'informations facilité par la fréquence des contacts, mais une perspective de partage où les implications de l'information sont explorées ensemble plus qu'échangées.

h) Le style de l'administrateur.

Pour qu'une information évaluative soit utilisée, il faut que les responsables du programme aient la capacité de l'intégrer dans la conduite des actions pour infléchir celles-ci.

Si un changement est suggéré par l'évaluation, il faut la volonté de la prendre en charge, d'accepter de nouvelles idées non par goût de l'innovation, mais par capacité à percevoir les bénéfices, à prendre des risques calculés.

Les relations entre les catégories

L'évaluation est un processus dynamique et le produit de multiples influences. Les différentes catégories doivent être vues dans leurs relations.

Au départ de l'évaluation, l'interaction entre les orientations des utilisateurs et l'approche de l'évaluateur va déterminer le plan de travail dans les contraintes existantes. Les attentes initiales définissent une première série d'options qui seront explorées et transformées. Ce travail préalable modifie en retour les attentes des utilisateurs. Les interactions se poursuivent durant l'évaluation et peuvent la transformer. Quand les informations deviennent plus abondantes, la forme de leur communication va jouer. Leur pertinence sera jugée ainsi que la crédibilité de l'évaluateur, les méthodes, pour les utilisateurs les mieux avertis, mais surtout la cohérence avec d'autres informations. Un dialogue sur les interprétations, les alternatives, améliore l'utilisation.

Les étapes vers le développement d'une théorie

Pour Alkin, Daillak et White, cette étude est une première étape. Il faut encore développer l'identification formelle des variables et de leurs interrelations ainsi que le lien avec l'utilisation.

Des études supplémentaires focalisées sur certaines caractéristiques sont à mener. Les méthodes devront aussi être adaptées comme l'usage de l'observation prospective et non plus seulement rétrospective.

4. Conclusions.

La conclusion d'Alkin et de ses collègues est que l'usage direct de l'évaluation existe, mais qu'il est rarement immédiat. Les données d'évaluation peuvent avoir une influence cumulative menant à des décisions majeures. Ils illustrent cela par trois des six cas analysés.

Le premier cas est très clair : une procédure de type expérimentale (programme bien délimité, groupe de contrôle, rapport d'évaluation écrit) aboutit à des conclusions négatives et le programme est abandonné.

Le second cas concerne un programme dont les résultats sont positifs, mais pour lequel la décision de financement est négative. Les résultats vont jouer un rôle important dans l'argumentation, mais c'est l'habileté politique à les utiliser par les partisans du programme qui permettra de modifier la décision de financement.

Le troisième cas est plus complexe encore. Des résultats médiocres et répétés aux tests de rendement vont déterminer la décision d'arrêt du financement fédéral du programme, mais après trois ans seulement. Cependant les données collectées sur le processus et servant initialement à l'ajustement du programme en cours vont, à l'inverse, avoir une influence majeure pour justifier la décision du district scolaire de maintenir son financement local.

L'utilisation moins directe est présente dans tous les cas analysés. L'utilisation secondaire de l'évaluation est fréquente, mais rarement de manière décisive.

L'évaluation est utilisée pour gérer le programme, pour individualiser l'enseignement. Elle influence le développement et les orientations. Elle est fonction de la fréquence de consultation de l'évaluateur, de l'implication de l'évaluateur dans les activités du programme.

Le contexte de "commande" de l'évaluation par des décideurs externes au programme est fort prégnant et limite l'utilisation des facteurs identifiés. Mais cette première liste de facteurs fournit des pistes de réflexion sur la complexité des facteurs contextuels et personnels qui peuvent intervenir non seulement dans l'utilisation, mais dans la conception et la mise en œuvre d'une évaluation. Les caractéristiques d'une pratique d'évaluation seront probablement fonction du contexte du programme et du rôle des utilisateurs, du contexte de l'évaluation et du rôle de l'évaluateur.

Mais le travail de Alkin et de ses collègues est surtout remarquable par sa méthodologie : rigueur dans la définition préalable de l'objet, rigueur dans l'analyse qualitative.

I.C. UNE REVUE DE LITTÉRATURE

La présentation des principaux résultats de la recherche sur l'utilisation des données de l'évaluation est surtout basée sur la revue de littérature de Cousins et Leithwood (1986). Et complétée par les données et les commentaires de trois autres articles de synthèse (Alkin, 1985, Alkin et Daillak, 1990, Leviton et Hughes, 1981). Nous proposerons tout d'abord une discussion sur les définitions du terme utilisation, puis un examen des méthodes adoptées par les recherches et, enfin, une synthèse des résultats de celles-ci dans le but de mettre en évidence les éléments qui seront utilisables dans le contexte de notre travail.

L'article de Cousins et Leithwood revoit la recherche empirique conduite durant la période de 1971 à 1985 sur l'usage des résultats d'évaluation à travers 65 études dans différents secteurs sociaux. Il en décrit les caractéristiques méthodologiques, les variables dépendantes et indépendantes et leurs relations.

La recherche de littérature s'est effectuée dans plusieurs banques de données sur les mots clés : *evaluation utilization, data use, test use, decision making, knowledge utilization*. L'élargissement à l'utilisation de données ou de systèmes de données de sciences sociales est justifié par leur support à la décision.

Un tableau classe les 65 études selon l'auteur, la taille de l'échantillon, la nature de celui-ci, le type de recherche, la nature des variables dépendantes et indépendantes, la théorie sous-jacente (42 % n'en mentionnent pas) et les instruments utilisés.

Les études d'évaluation sont définies comme le résultat de l'utilisation des méthodes des sciences sociales pour répondre aux questions des décideurs. Le rôle de l'évaluateur est de mener l'étude d'évaluation et de communiquer les résultats aux décideurs. Les résultats d'évaluation sont définis largement comme les informations associées aux produits de l'évaluation : données, interprétations, recommandations, pendant ou à l'issue de l'évaluation. L'objet de l'évaluation peut être multiple : curriculum, programme, projet, école, cours, personnel ou étudiant.

Une structure d'interprétation est proposée qui liste 12 facteurs répartis en deux groupes : ceux qui sont associés aux caractéristiques de l'évaluation et ceux qui ressortissent du contexte de la décision ou de la situation politique.

L'objet de la revue de littérature est à la fois trop large et trop étroit en regard de nos propres critères. Il est trop large parce qu'il ne fait pas la distinction entre l'évaluation et la recherche. Il est trop étroit parce qu'il réduit le produit de l'évaluation aux informations fournies par celle-ci. Néanmoins, cet article offre un cadre d'analyse pertinent car il inclut les études sur l'utilisation de l'évaluation.

1. Les définitions de l'utilisation.

Dans cette revue, l'utilisation (ou l'usage) est la variable dépendante. Les facteurs d'influence sont les variables indépendantes. De nombreuses taxonomies sont disponibles.

Alkin (1985) identifie une cinquantaine de facteurs associés à l'utilisation, qu'il regroupe en quatre catégories : les caractéristiques de l'évaluation et celles de l'évaluateur, les caractéristiques de l'utilisateur et celles de la situation de décision.

Cousins et Leithwood (1986) considèrent 12 variables indépendantes regroupées en deux grandes catégories qui suivent la même division : l'implantation de l'évaluation et la situation de décision. Ce sont ces deux catégories qui serviront pour présenter les résultats.

Leviton et Hughes (1981) proposent souvent les mêmes variables, mais regroupées différemment : la pertinence, la communication, la prise d'information, la crédibilité et l'implication.

La classification des variables dépendantes par Cousins et Leithwood (1986).

Cousins et Leithwood (1986) envisagent trois types de variables dépendantes : la prise de décision (usage instrumental), l'apprentissage (usage conceptuel) et le processus de prise d'informations. Certaines études ne clarifient pas le type d'utilisation. Les auteurs les ont regroupées dans une catégorie "usage potentiel".

a) Sur le plan instrumental, on trouve diverses catégories :

- décisions sur le financement du programme (renouvellement, changement, financement initial) ;
- décisions sur les actions du programme ;
- décisions sur la gestion du programme.

b) Sur le plan conceptuel, l'évaluation influence de manière indirecte la représentation idéale du programme :

- les concepts généraux de l'équipe de gestion ;
- la confirmation d'impressions préalables ;
- l'amélioration de la confiance, l'identification des forces et des faiblesses.

c) Le plan de la prise d'information est abordé par diverses études : demandes supplémentaires d'un rapport d'évaluation, souvenirs des recommandations, réactions aux aspects formels d'une évaluation, utilisation d'un système d'information managériale, compréhension des données d'évaluation, compréhension d'un rapport d'évaluation.

d) Le plan "potentiel" comprend les variables liées à l'utilisation, comme la satisfaction vis-à-vis de l'information reçue, ou l'estimation de son influence sur les décisions, le degré d'accord avec les recommandations.

Cet indicateur suppose que le décideur est le meilleur juge de l'efficacité du programme et que l'acceptation des recommandations est un bon critère prédictif de l'usage.

L'ambiguïté du concept d'utilisation est donc clairement perceptible au travers de la grande variabilité qui marque la définition de la variable dépendante.

Quel rapport peut-on faire avec d'autres auteurs ?

Patton (1978) partait du point de vue de l'utilisateur (usage potentiel), mais distinguait, au niveau des résultats, un usage direct (instrumental) et un usage indirect (conceptuel). Alkin, Daillak et White (1979) arrivaient au même point en distinguant la prise de décision et l'attitude vis-à-vis du programme et en insistant sur le caractère cumulatif des deux aspects.

Leviton et Hughes (1981) consacrent une partie de leur revue de littérature au problème de la définition de l'utilisation.

Pour ces derniers, le terme utilisation est limité à l'usage des résultats pour des programmes et des politiques. Il ne concerne pas l'utilisation par les universitaires ou la presse.

a) Leviton et Hughes énoncent tout d'abord une série de distinctions.

- Une distinction entre processus d'information et modification dans la façon de penser et d'agir.
 - Le processus d'information concerne le transfert ou la tentative de transfert des conclusions dans l'action. C'est le processus d'information que Weiss est la première à identifier et que l'on retrouve chez Cousins et Leithwood. Leviton et Hughes précisent que la lecture d'un rapport n'est pas une "utilisation".
 - La modification dans la façon de penser et d'agir implique un changement dans les activités, dans les attitudes, les croyances vis-à-vis d'un problème, d'une activité.
- Une distinction aussi entre utilité, utilisation et impact.
 - L'utilité est la pertinence des conclusions pour les problèmes du programme ou de la politique. C'est la non utilisation d'évaluation jugée potentiellement utile qui a motivé les études sur l'utilisation de l'évaluation.
 - L'utilisation est l'application des informations fournies par l'évaluation.

- L'impact est la modification effective d'un programme ou d'une politique par les conclusions de l'évaluation. L'impact peut être faible malgré l'existence de tentatives d'utilisation.

L'impact, l'utilisation et l'utilité sont trois concepts liés mais distincts. Une évaluation peut être de faible utilité parce qu'elle est de mauvaise qualité ou que ses implications ne sont pas praticables. Elle peut être inutilisée parce que l'utilité n'est pas reconnue ou qu'on choisit de ne pas l'utiliser. Elle peut ne pas avoir d'impact parce qu'on choisit de ne pas l'utiliser ou parce que de nombreux facteurs contraignants interviennent.

- Alkin et Daillak (1990) font la même distinction entre utilisation et impact.
 - L'impact attendu d'une évaluation doit prendre la forme d'une meilleure décision sur les programmes et les politiques. Plus largement, l'impact d'une évaluation est l'influence discernable sur les activités ou attitudes des individus ou des groupes.
 - L'utilisation doit être distinguée de l'impact. Elle concerne l'application intentionnelle d'une information ou d'un processus d'évaluation. Certains impacts sont dus à une utilisation prévue, mais d'autres impacts inattendus peuvent apparaître. Par cette remarque, ces deux auteurs posent en quelque sorte le problème de l'évaluation comme celui d'une action en soi, qui a des objectifs, des effets attendus ou inattendus.

b) Leviton et Hughes proposent ensuite des catégories d'utilisation.

Les contextes de l'évaluation définissent différentes utilisations, que l'on peut catégoriser en instrumentale, conceptuelle et persuasive (ou symbolique).

- Usage instrumental
L'utilisation est directe, observable, par exemple la décision de financement d'un programme sur la base d'une évaluation d'efficacité.
- Usage conceptuel
L'utilisation n'est pas directement observable. Elle porte sur la manière de penser, de juger un programme.
- Usage persuasif ou symbolique
L'utilisation suppose une influence interpersonnelle. L'évaluation sert à supporter ou à affaiblir une position politique.

La définition de l'usage instrumental et de l'usage conceptuel pose problème. Il ne faut pas confondre les concepts "instrumental" et "conceptuel" avec les modalités d'opérationnalisation de ceux-ci. Un usage instrumental peut rester confidentiel, non repérable dans des écrits tandis qu'un usage conceptuel peut être identifié dans des textes de référence sans mener à des changements concrets. De plus, la frontière entre usage conceptuel et instrumental est parfois minime et graduelle.

Au terme de leur revue de littérature, Leviton et Hughes concluent qu'il n'y a pas d'usage spécifiquement instrumental ou conceptuel, mais de multiples usages en fonction d'objectifs variables dans le temps. Ils proposent de déterminer l'usage en fonction des cycles de décision bureaucratique ou politique (rédaction des projets de loi, examen devant les assemblées, surveillance des programmes, ...). Cependant, il n'est pas toujours facile de déterminer les étapes de ces cycles.

Alkin et Daillak (1990) proposent une autre distinction plus subtile entre deux formes d'impact. L'impact de l'information est l'impact produit par les conclusions et les recommandations. L'impact du processus d'évaluation est dû au processus même d'évaluation qui peut influencer les activités et les attitudes en cours de programme.

- L'impact de l'information

Alkin envisage divers modes d'influence des informations évaluatives.

Il y a tout d'abord une influence directe sur la prise de décision. Mais cette influence peut être importante ou partielle, visible ou invisible. L'influence partielle parmi plusieurs sources d'information est le cas le plus fréquent. L'évaluation a une influence cachée lorsqu'on ne peut l'associer précisément aux effets observés. C'est le cas le plus fréquent dans les systèmes ou programmes qui réussissent bien ou dans le cadre d'actions routinières. Le succès d'un programme ou d'une action peut être soutenu par les résultats de l'évaluation qui agissent comme un feedback positif, mais sans signes extérieurs.

La deuxième modalité importante est l'influence sur les attitudes et croyances, parfois appelée influence "conceptuelle". Ce type d'influence peut être important, sans toutefois déboucher sur des actions, ou alors avec un effet de retard très marqué. C'est ainsi que les résultats positifs exercent une influence pour conforter des positions existantes dans le cadre de pression externe (débat politique, par exemple) ou pour valider des prévisions, des intuitions.

- L'impact du processus

Les processus d'évaluation peuvent influencer une situation éducative de plusieurs manières.

- De manière évidente en consommant des ressources humaines, matérielles, financières et temporelles.
- De manière plus subtile en modifiant le contexte éducatif lui-même. Les objectifs évalués peuvent être différents des pratiques réelles et devenir des priorités dans l'investissement pédagogique. L'évaluation peut provoquer des changements dans les modes de communication, en impliquant par exemple, des échanges, là où il n'y en avait pas. L'évaluation peut provoquer de l'anxiété.
- Enfin, le processus d'évaluation peut aussi avoir une signification symbolique. La seule existence d'une évaluation est interprétée comme un signe de gestion, de responsabilité, de justification pour le public ou pour le législateur. Elle légitime le programme.

Alkin et Daillak élargissent ainsi considérablement le rôle de l'évaluation au-delà de la communication d'une information et permettent de confronter les propositions précédentes à notre propre grille d'analyse.

Discussion des concepts de l'utilisation à la lumière de la grille APEP.

Le modèle de l'utilisation de l'évaluation basé sur un processus de communication d'une information entre l'évaluateur - émetteur et le décideur - récepteur est manifestement incapable de rendre compte de toutes les facettes des relations entre l'évaluation et l'action, même si l'on multiplie les étapes : prise d'information, usage potentiel (appropriation de l'information perçue à travers la satisfaction, le degré d'accord), usage conceptuel, puis usage persuasif et enfin usage instrumental.

Il faut donc concevoir l'usage de l'évaluation plus largement comme un processus au cœur de l'action, qui influence l'action dès le départ ; comme un processus le plus souvent invisible ; comme un processus qui agit par sa méthodologie, le jugement qu'il produit et les décisions qu'il entraîne ; comme un processus qui modifie les rapports entre personnes et organisations.

La vision plus large de l'évaluation que nous avaient donnée les auteurs comme Patton, Alkin ou Monnier et bien d'autres fait de l'évaluation une pratique sociale, complexe, en interaction avec le processus d'action. La tentative de définition de l'utilisation de l'évaluation renvoie donc de manière plus large vers une définition de l'évaluation.

De la définition de l'utilisation, nous avons conservé la composante "décision" en distinguant les formes énoncées, visibles et en identifiant son caractère interne ou externe, informel ou formel. Le caractère formel ou informel ne correspond pas à une distinction conceptuelle versus instrumentale, mais à un usage instrumental direct, immédiat par rapport à un usage indirect, différé. Elle correspond à la distinction de Alkin, Daillak et White (1979) entre une utilisation dominante ou cumulative dans un processus de décision et à la distinction de Sheerens (1990) entre un processus rationnel ou cumulatif. La prise d'information, l'influence potentielle ou conceptuelle sont assimilées à la dimension implicite, car elles recouvrent des processus intermédiaires non observables. La fonction symbolique n'est pas considérée, car il s'agit le plus souvent d'un jugement de valeur des auteurs sur la portée politique ou non de l'utilisation.

2. Méthodologie.

Cousins et Leithwood (1986) proposent un inventaire des méthodologies utilisées dans les 65 études recensées. Leviton et Hughes (1981), puis Alkin et Daillak (1990) examinent les difficultés méthodologiques de la recherche sur l'utilisation.

- a) Selon Cousins et Leithwood, on trouve trois types de recherche : rétrospective, longitudinale et expérimentale.

La recherche rétrospective.

Ces études sont centrées sur des évaluations passées et travaillent sur des données qualitatives issues principalement de la mémoire de décideurs, de commanditaires et d'évaluateurs. La fiabilité est limitée et cette méthodologie exclut la manipulation des variables indépendantes.

Une majorité des études sont de type rétrospectif (35). Il y en a 17 qui sont des études de cas, les autres sont des surveys. Cinq de ces 18 surveys ne nous intéressent pas car ils concernent des études sur l'usage des tests standardisés.

Les surveys utilisent les instruments dirigés : questionnaires, entretiens dirigés par téléphone. La majorité concernent le milieu scolaire, mais on trouve des études en milieu politique (2) et en milieu social (1). Il n'y a pas de référence à des théories.

Les 17 analyses de cas utilisent des méthodes plus approfondies : entretiens, analyse de documents ou observation participante. Trois études viennent de Alkin, mais d'autres sont prises en charge par le centre d'étude de l'évaluation de l'Université de Californie dont il est le directeur. Le nombre de cas varie de 1 à 273. Les évaluations ont été mises en œuvre par les programmes eux-mêmes, par des centres d'évaluation externes ou, le plus souvent, par des évaluateurs individuels. Des modèles théoriques sont plus souvent cités. Les personnes étudiées sont variées : décideurs, évaluateurs, politiciens. Quatre études, dont deux en santé mentale, concernent un autre milieu que l'école.

La recherche longitudinale.

Elle examine l'influence de données rassemblées avant, pendant ou après l'implantation de l'évaluation. La manipulation des variables indépendantes est parfois réalisée, mais la durée de ces études est trop limitée pour enregistrer l'impact des résultats de l'évaluation.

Sur les 9 études longitudinales, l'une est expérimentale, mais elle concerne l'usage de tests. Les 8 autres sont des études de terrain. Quatre concernent le domaine de la santé mentale communautaire. Elles étudient l'influence de résultats d'évaluation sur des directeurs de programmes, des coordonnateurs, des administrateurs et des membres du personnel. Les autres étudient des cas isolés en recourant à des techniques d'observation participante.

La recherche expérimentale.

Elle emploie habituellement des rapports simulés pour tester dans des conditions contrôlées les effets de certaines variables indépendantes. Mais la simulation ne peut reproduire la plupart des conditions des situations de décision.

Il y a 21 études de ce type, dont 19 viennent de la même équipe (Brown, Newman et Braskamp, cité par Cousins et Leithwood, 1986) et utilisent la théorie de la communication (15) ou de la décision (4) sur des étudiants (8) des enseignants (10) ou des directeurs (1).

b) Leviton et Hughes comme Alkin et Daillak dressent un tableau très mitigé des résultats des recherches sur l'utilisation de l'évaluation.

Ils soulèvent surtout des problèmes méthodologiques.

Il y a d'abord des difficultés générales dues à l'objet de l'étude et à sa définition (Alkin et Daillak, 1990).

La mesure de l'impact de l'évaluation se heurte à quatre types de difficultés : le manque de visibilité de l'impact, la présence de nombreuses autres influences, les problèmes de diffusion de l'information évaluative et la variété des impacts.

La formation des choix, des décisions, des attitudes est une activité mentale qui laisse des traces très variables. Le jugement individuel est souvent moins documenté que le jugement élaboré par des groupes. Le maintien d'une activité laisse rarement des traces écrites. Les mesures d'impact sont souvent fondées sur les déclarations des décideurs ou inférées d'actions mises immédiatement en route par ceux-ci après une information évaluative.

Les appréciations des décideurs et des participants au programme sont peut-être le seul repère disponible, mais elles sont aussi sujettes à erreurs, car ceux-ci font difficilement la part de la contribution de l'évaluation parmi d'autres influences.

Les difficultés de mesure sont aussi fonction du temps écoulé, qui augmente la possibilité d'autres influences et qui permet une diffusion de l'information dans le programme ou le système politique. L'origine de l'information est rapidement perdue. Les travaux de Keeves (1990) montrent que l'information cognitive ne diffuse pas de la même façon qu'une innovation technologique. Le plus souvent, elle influence la société de manière implicite et diffuse.

Enfin, l'impact de l'évaluation, de par son processus et ses produits, est multiple et varié au sein d'un système social, ce qui rend les mesures uniques inadéquates. Une même décision peut avoir des origines multiples, comme une même évaluation peut entraîner des décisions multiples ou des évocations multiples.

Il y a ensuite des difficultés propres aux méthodes de recherche (Leviton et Hughes, 1981).

Les problèmes méthodologiques abondent : définitions imprécises de l'utilisation, types de méthodologie.

Les conclusions des recherches sont fonction de la définition de l'évaluation et de l'utilisation, du type de personnes interrogées (l'utilisation est différente selon les responsabilités) et du contenu, du thème des programmes.

Les auteurs considèrent l'utilisation de la recherche appliquée, l'utilisation de projets de recherche ou d'évaluation particuliers, les réactions à la lecture d'une évaluation. Ils étudient des groupes hétérogènes, des administrateurs seuls, des praticiens, des équipes de politique. Les programmes portent sur des problèmes éducatifs, sociaux.

Les analyses de cas comme celles de Alkin sont les plus nombreuses. Elles ont comme point fort une meilleure probabilité de mettre au jour des liens entre les conclusions de l'évaluation et l'évidence de l'utilisation. Leur faiblesse provient d'un manque de généralisation et du poids du contexte. Dans les études rétrospectives, le laps de temps entre l'utilisation et l'étude peut être élevé, entraînant de nombreux biais (oublis, incident critique exceptionnel, ...). Le petit nombre de cas entraîne des difficultés de comparaison. Une voie de développement est l'analyse d'un échantillon de cas représentatifs des évaluations dans un secteur d'activités.

3. Synthèse des résultats.

Cousins et Leithwood proposent une analyse systématique des relations entre les variables et les quatre types d'usage qu'ils ont identifiés (instrumental, conceptuel, informatif et potentiel).

Ils calculent un index de prévalence (ou rang) : E/CxO est le nombre d'études (E) sur le nombre de citations (C) multiplié par le nombre de relations observées (O). Cet index permet de comparer la force relative entre les variables, à la manière des méta-analyses.

- Le nombre de citations (C) est l'ensemble des relations observées et non observées.
- Une relation observée (O) signifie une relation statistique dans les études quantitatives et une relation relatée dans les études qualitatives.
- La composante E/C est un facteur d'ajustement entre des études de nature différente, l'observation d'absence de relation étant typique des démarches quantitatives et n'étant pas mentionnées en général dans les démarches qualitatives.

La "qualité" de l'évaluation est le facteur qui apparaît le plus souvent dans les études quel que soit le type d'utilisation. Mais c'est une variable peu précise et qui donne des résultats contradictoires. Elle ne semble pas devoir être retenue pour notre étude comme un élément descriptif majeur d'une pratique d'évaluation.

Les facteurs d'influence de l'utilisation instrumentale et conceptuelle sont nettement différents entre eux. C'est la première qui retiendra notre attention, mais sans la lier à une utilisation directe ou immédiate. Nous préférons retenir comme indicateur les formes énoncées socialement.

L'utilisation instrumentale semble surtout influencée par l'adéquation des réponses apportées (conclusions) aux attentes préalables des utilisateurs et ensuite par leur pertinence par rapport aux problèmes rencontrés. Donner les réponses attendues (rang 2) influence plus que donner les réponses aux questions posées (rang 4). Mais répondre aux questions de l'utilisateur vaut mieux pour l'évaluateur que de les fixer lui-même, du moins s'il vise une utilisation instrumentale. Ces questions doivent porter surtout sur des problèmes immédiats, concrets, liés au contexte (rang 3). L'évaluateur interne, ou bien au fait des problèmes du programme, est plus efficace en regard de l'utilisation. L'engagement de l'évaluateur et l'implication de l'utilisateur (rang 6) seront ensuite déterminants. La présence d'informations concurrentes (rang 5) peut aussi perturber l'utilisation. La forme de la communication et les délais ont moins d'importance.

L'utilisation conceptuelle est liée différemment aux diverses variables. C'est l'implication de l'utilisateur (rang 2), son intérêt initial qui entrent le plus en jeu. Puis viennent le climat politique (rang 4), les caractéristiques générales personnelles (rang 4) et la présence d'autres sources d'information (rang 4). La pertinence est au dernier rang de l'influence vis-à-vis de l'utilisation conceptuelle. L'utilisateur n'aurait pas de questions précises, de décisions spécifiques. Ce serait en fonction de ses intérêts personnels et professionnels que l'utilisateur potentiel recourrait aux données de l'évaluation.

La distinction instrumental / conceptuel semble recouvrir les deux types d'études répertoriés par Cousins et Leithwood : l'utilisation de données provenant de l'évaluation d'une part et de la recherche, d'autre part. Ces deux auteurs concluent en suggérant un modèle d'intervention en évaluation proche des démarches "orientées vers le client" proposées par Patton et Alkin. Il faut chercher des informations utiles pour les décisions, impliquer les utilisateurs pour assurer la pertinence, assurer l'interactivité à toutes les étapes pour résoudre les problèmes et créer un climat de confiance.

Alkin et Daillak (1990) en font aussi leur conclusion.

De nombreux facteurs semblent à la portée de l'évaluateur pour augmenter l'utilisation et l'impact de l'évaluation : la pertinence, la crédibilité, la manière de partager les informations entre évaluateur et utilisateurs, les styles de traitement de l'information des administrateurs, le degré d'implication des utilisateurs potentiels dans l'évaluation.

Le partage de l'information facilite l'impact de plusieurs manières : au départ en ciblant les intérêts et les besoins en information ; en cours d'évaluation pour répondre à l'évolution des besoins ; à la fin, par une diffusion soignée d'informations précises et compréhensibles avec un suivi et un soutien aux utilisateurs.

Un point semble acquis, la composante personnelle et psychologique de l'utilisation de l'information.

Cousins et Leithwood prennent cependant la précaution de soulever la question de l'objectivité de l'évaluateur. Centrer les évaluations sur les perceptions des utilisateurs peut sembler restrictif et ne pas servir d'autres besoins "supérieurs". L'engagement dans le processus peut aussi sembler mener vers la cooptation, la perte d'objectivité, et une menace sur l'intégrité de l'évaluateur. Ces inconvénients doivent être pesés par rapport au coût d'une non-utilisation.

L'association de l'utilisateur au processus d'évaluation est un équilibre entre le non-usage et le copinage.

Alkin et Daillak (1990) permettent de replacer la recherche sur l'utilisation de l'évaluation dans un cadre plus large.

La nature de la prise de décision dans les organisations a une implication importante sur l'évaluation. Les décisions explicites, identifiables à un moment donné semblent relativement peu fréquentes. Le processus de décision est plutôt cumulatif, résultant de choix d'importance variable par différents acteurs qui n'ont pas nécessairement conscience d'être engagés dans une décision majeure sur le plan du programme. Ceux-ci basent leur action sur des procédures routinières relevant de considérations immédiates et l'évaluation leur apparaît inutile par rapport au travail quotidien.

Quand une décision formelle doit être prise, elle repose sur un nombre limité d'alternatives et ne prend en compte que quelques données pour chacune d'entre elles plutôt que de reposer sur une recherche exhaustive des options. Ces données, ces options sont souvent de nature politique et bureaucratique et ne rencontrent pas les options des évaluateurs - chercheurs orientés vers le recueil de résultats généralisables. L'évaluation ne rencontre dès lors qu'une partie des besoins des décideurs.

Alkin et Daillak acceptent donc l'idée d'une évaluation interne, mais en la limitant à la régulation des actions routinières.

4. Conclusions.

Le bilan de la recherche sur l'utilisation de l'évaluation confirme l'image d'une répartition figée des rôles entre diverses évaluations.

- Il y aurait une évaluation - recherche ou une recherche évaluative qui poursuit ses propres buts et qui enrichit les connaissances disponibles sur l'action. Les utilisateurs potentiels, les décideurs utilisent ce patrimoine conceptuel commun en fonction de leurs intérêts et besoins.
- Il y aurait une évaluation "sommative" chargée de fournir des informations sur l'efficacité des programmes. Mais ces données ne sont que rarement utilisées de manière directe. Elles sont un élément parmi d'autres qui fondent le jugement et la décision de l'administrateur. Cette mode d'évaluation sert surtout à conforter des décisions.

- Il y aurait une évaluation orientée vers la décision pour répondre à des questions précises, concrètes. Son utilisation dépend de l'interaction entre l'évaluateur et le décideur, mais elle serait aussi fonction des réponses attendues.
- Il y aurait enfin une gestion interne, routinière, peu intéressante et qui, en fait, n'a besoin ni d'évaluation, ni d'évaluateur.

Ce tour d'horizon de la recherche sur l'utilisation de l'évaluation met en évidence les différentes dimensions nécessaires à notre analyse.

- Les trois composantes de l'évaluation apparaissent en filigrane de la définition des objets d'études et de leur interprétation.
- L'évaluation comme pratique sociale, objet d'enjeux multiples, est omniprésente dans toutes les variables identifiées. C'est par le concept d'explicitation que nous tentons d'en rendre compte.
- La position interne et externe de l'évaluateur est un élément d'analyse, mais qui est peu exploité, dans la mesure où les études ne prennent en compte que des évaluateurs professionnels.
- Le degré de formalisation n'est envisagé que du point de vue de la qualité technique. Il est souvent limité au processus méthodologique de recueil des informations.

C'est l'hypothèse d'une unité de nature du processus d'évaluation au sein de l'action qui conduit notre travail. L'évaluation jouerait le même rôle dans le cadre d'une action courte et limitée au sein d'un programme ou d'une action longue et complexe dans le cadre d'une politique. C'est le type de contexte qui donnerait à l'évaluation des fonctions différentes, en mettant l'accent sur certaines composantes et certains acteurs. Cette revue de littérature met également en perspective un champ de recherche jusqu'ici négligé : celui de la description des pratiques internes de l'évaluation. Celles qui seraient "invisibles".

Les recherches menées pour décrire les pratiques occuperont la deuxième partie de ce chapitre.

II. LA PRATIQUE DE L'EVALUATION

Plus que les recherches sur l'utilisation de l'évaluation, c'est l'étude du processus d'évaluation qui enrichirait notre sujet d'étude. Malheureusement, nous n'avons pu identifier que des tentatives isolées d'analyse de ce sujet. Ces quelques études seront plutôt proposées à titre d'exemple méthodologique que pour les informations qu'elles nous apportent.

Les deux premiers exemples (Alkin et Ellett, 1979, Williams, 1989) s'interrogent sur la genèse et l'utilisation des modèles théoriques. Ces modèles sont des théorisations de pratiques professionnelles. L'étude de cette théorisation et de l'application de ces "théories" apportent des informations sur les pratiques des professionnels.

Les exemples suivants (Marsch, Newman et Bover, 1981, Nioche, 1982, Sensi, 1994) sont des enquêtes sur des pratiques d'évaluation. La dernière étude (Shadish et Epstein, 1987) s'intéresse à la pratique des évaluateurs.

II.A. L'ETUDE DES THEORIES

Nous avons longuement décrit de nombreux modèles théoriques en cherchant à les comparer pour comprendre ce qu'est l'évaluation et vérifier la pertinence de notre grille d'analyse. Certains auteurs ont entrepris des études très spécifiques en prenant ces modèles comme des objets d'étude.

1. Alkin et Ellett (1979) : une analyse naturaliste.

Dans cet article, Alkin et Ellett s'intéressent aux stratégies que les auteurs utilisent pour développer leurs théories et non au contenu des théories elles-mêmes. Ces auteurs ne catégorisent plus les modèles théoriques. Ils étudient le processus de théorisation.

Nous avons retenu cette démarche comme un exemple d'une analyse particulière des pratiques d'évaluation. Cet angle de vision est intéressant, car il cherche à mieux comprendre la nature de différentes théories particulières. Il peut être utile s'il permet de saisir l'apport respectif de l'expérience de l'auteur, de ses principes théoriques sous-jacents et, le cas échéant, de son travail d'expérimentation.

Les théories qui apparaissent souvent comme des énoncés solides sont en fait des tentatives, des essais de clarification. Ce caractère provisoire est généralement oublié par le lecteur. Le développement d'une théorie est une entreprise humaine où les facteurs personnels jouent un rôle majeur.

Le but de Alkin et Ellett est de considérer les aspects personnels et informels du développement d'une théorie à partir d'un petit nombre de cas. Leur travail cherche à montrer les stratégies suivies, les étapes, les éléments déclencheurs et les freins de ces constructions théoriques.

Ils choisissent une méthodologie naturaliste, qualitative. Cette méthodologie est originale. Elle s'appuie sur l'interprétation des idées d'un auteur par des étudiants uniquement sur la base des écrits de celui-ci. Ces travaux sont envoyés à l'auteur en lui demandant d'y réagir. C'est l'analyse de cette réaction et de son argumentation qui permet l'étude du processus de théorisation. Trois auteurs sont analysés de cette manière : Rippey (l'évaluation transactionnelle), Guba (l'évaluation naturaliste) et Wolf (l'évaluation judiciaire).

La réaction des auteurs montre qu'ils ne sont pas toujours conscients de leurs propres évolutions, de certaines ruptures. La confrontation avec la reconstruction leur permet de redéfinir, clarifier, réinterpréter leurs énoncés.

L'analyse de Alkin et Ellett fait apparaître la dynamique du processus de théorisation. Elle attire l'attention sur la dimension historique, évolutive des pratiques d'évaluation et de leur théorisation. Les théories de l'évaluation sont plus des pratiques théorisées qu'un ensemble de connaissances vérifiées par la recherche. Elles sont relativement sensibles aux contextes qui les ont vues naître et se développer. Elles n'ont pas subi le processus de "purification" typique des démarches scientifiques généralisantes.

A la lecture de cette étude, il semble légitime d'affirmer que la pratique évaluative reste largement à définir et à préciser.

2. Williams (1989) : une analyse numérique.

Une autre tentative d'analyse des modèles théoriques est celle de Williams qui étudie de manière quantitative les différences entre les modèles théoriques et entre les pratiques de ces modèles. Elle comporte deux phases : une mesure et une classification.

a) Méthodologie.

Quatorze leaders de l'évaluation ont répondu à un questionnaire sur la similarité de leur théorie avec celles d'autres évaluateurs et sur les caractéristiques de leurs pratiques actuelles. Ces quatorze personnes sont reconnues comme des théoriciens de l'évaluation et considérées comme particulièrement au fait des positions théoriques existantes. Il s'agit de Abt, Alkin, Cook, Cronbach, Eisner, Freeman, House, Patton, Popham, Scriven, Stake, Stufflebeam, Weiss et Wholey.

Les réponses ont été traitées par la technique des échelles multidimensionnelles pour fournir des cartes sur les dimensions sous-jacentes aux théories et pratiques. Les résultats indiquent que les différences perçues entre les théories sont plus importantes et plus complexes que les différences perçues entre les pratiques.

Phase 1 : la mesure

Le questionnaire adressé comprend deux pages et occupe de dix à vingt minutes. Il comprend deux questions. La première demande de situer la similarité globale de la position théorique de la personne interrogée avec chacune des autres. La seconde demande à l'auteur de situer sa pratique actuelle par rapport à des concepts présélectionnés.

Question sur la théorie.

"Quelle est la place générale de votre vision de l'évaluation comparée aux vues des autres évaluateurs ? Répondez pour chaque évaluateur en plaçant une croix sur l'échelle ci-dessous sur la dimension "semblable - dissemblable"."

L'échelle proposée est de neuf centimètres. Chaque théoricien est donc situé vis-à-vis de chaque autre sur une échelle "semblable - dissemblable". La distance considérée est la moyenne de celle exprimée par deux scores. Par exemple, Scriven se situe vis-à-vis de Cronbach et Cronbach se situe par rapport à Scriven. La diagonale de cette matrice est zéro. Cronbach est semblable à Cronbach.

Question sur la pratique.

"Comment situeriez-vous votre approche de l'évaluation aujourd'hui en regard de chacune des dimensions suivantes ? Placez une croix sur le point de l'échelle."

Les sept échelles proposées sont :

- *mesures numériques >< attributions descriptives / interprétatives ;*
- *affirmations causales >< affirmations descriptives ;*
- *participation des acteurs >< non participation des acteurs ;*
- *publics spécifiés >< audience générale ;*
- *juge l'efficacité >< ne juge pas l'efficacité ;*
- *juge la valeur des résultats >< ne juge pas la valeur des résultats ;*
- *s'accommode de la situation politique >< ignore la situation politique.*

Ces sept dimensions sont suggérées par la littérature (House, 1986, Shaddish et Epstein, 1987). Chaque échelle fait neuf centimètres. La matrice obtenue pour chaque auteur est comparée à celle de chacun des autres auteurs.

Phase 2 : la classification

Une technique statistique d'analyse d'échelles multidimensionnelles est appliquée pour localiser la position de chaque auteur sur une carte.

b) Résultats.

Approche théorique

La carte théorique met en évidence quatre dimensions. Chaque auteur est classé selon son degré de convergence avec chaque dimension. La qualification de chaque dimension est déterminée en cherchant les caractéristiques théoriques qui représentent le mieux les différences entre les évaluateurs placés à chaque bout des échelles.

- La première dimension est nommée "*Qualitatif versus Quantitatif*". Les évaluateurs aux extrémités de l'échelle sont Eisner et Stake pour le pôle qualitatif, Freeman et Cook pour le pôle quantitatif.
- La deuxième dimension est nommée "*Justification de la valeur versus Orientation politique*". Scriven et Popham sont des partisans de la production d'une information qui établit la valeur d'un programme particulier. Cronbach et Stake prônent quant à eux la production d'une information pour guider les décisions politiques.
- La troisième dimension oppose Patton et Weiss d'un côté, Abt et Eisner de l'autre, sur l'axe "*Participation versus Non-participation du client*".
- La quatrième dimension fait apparaître une différenciation entre Wholey et Cook d'une part et Alkin et Freeman d'autre part sur l'axe "*Utilisation générale versus Utilisation pour la décision*".

Les dimensions 1 et 2 représentent 65 % de la variance des distances entre les théories. Les dimensions 3 et 4 comptent pour 17 % supplémentaires. Les deux premières dimensions sont donc celles qui séparent le plus les évaluateurs entre eux.

Approche pratique

Deux dimensions ressortent de l'analyse des échelles sur l'approche pratique.

La première dimension regroupe les items 1 et 2 en opposant "*interprétation et présentation descriptive*" d'une part et "*mesures numériques et liens causal*" de l'autre. Stake, Eisner et Wholey sont situés sur le pôle qualitatif et Abt, Popham, Freeman, Scriven et Cook sont sur l'autre pôle plus quantitatif. Wholey surtout et Weiss obtiennent des positions différentes de celles de la première dimension théorique en se trouvant dans le pôle plus descriptif et interprétatif. Scriven a aussi une position quantitative plus marquée.

La deuxième dimension se centre sur l'item 4 appelé "*Public général versus Utilisateur spécifié*". Cronbach, Scriven et Eisner représentent le pôle grand public ; Wholey et Abt représentent l'autre pôle.

L'existence de deux dimensions permet de regrouper les auteurs en fonction de la combinaison de leur position sur celles-ci : la méthodologie et l'utilisation. Deux regroupements majeurs apparaissent. Dans le premier groupe, on trouve les partisans du qualitatif et de l'usage direct avec Alkin, House, Patton, Stake, Stufflebeam, Weiss et Wholey. Dans le second groupe, les tenants de l'approche quantitative à usage indirect avec Abt, Cook, Freeman et Popham.

Comparaison des approches pratique et théorique

La comparaison entre les réponses aux deux questions indique que les différences sont plus grandes vis-à-vis de la question sur les théories que de celle sur les pratiques. La carte théorique est plus complexe (quatre dimensions) que la carte pratique (deux dimensions) et montre des différences individuelles plus fortes. Selon Williams, ces différences tiennent plus à la complexité des approches théoriques qu'à la nature des pratiques individuelles des auteurs.

c) Discussion.

L'étude de Williams comporte certaines limites. Elle s'appuie sur le jugement subjectif des auteurs interrogés. L'analyse numérique permet de raffiner et de cristalliser les opinions des auteurs, mais ne dit rien sur leurs pratiques réelles.

A cet égard, la distinction que fait Williams entre théorie et pratique est artificielle. Les auteurs se situent d'une part entre eux sur base de leur perception des autres modèles (approche théorique). D'autre part, ils se situent tout aussi subjectivement sur des dimensions définies au préalable par Williams sur la base de sa propre lecture de la littérature (approche dite "pratique" mais qui n'est peut-être que leur position théorique actuelle, car la question reste assez vague). Williams ne pose pas la question d'une différence entre actes et paroles chez ces auteurs ; la différence de complexité entre les images reflétées par la première et la deuxième question semble induite par la forme de la question plus que par une différence théorie / pratique.

Malgré tout, l'intérêt de l'étude de Williams est de renforcer l'image d'une approche plurielle de l'évaluation à travers la diversité de ses modèles théoriques, à travers les positions des évaluateurs professionnels. Cette démarche met en évidence l'existence de continuums différenciant les évaluateurs sur des options importantes déjà traitées dans les chapitres précédents, en particulier sur la dimension "méthode" et sur la dimension "utilisation".

En parallèle, les travaux de Alkin sur l'utilisation montrent en effet que les praticiens n'utilisent pas les modèles théoriques en tant que tels, que les schémas les plus rigoureux sont très rares. Une étude des pratiques reste donc à faire.

II.B. L'ETUDE DES PRATIQUES D'EVALUATION

1. Une description des pratiques d'évaluation des politiques de la commission européenne.

Sensi (1994) réalise une analyse comparative des évaluations des programmes d'éducation à la Commission européenne.

a) Définition de l'évaluation.

Sensi propose quatre dimensions d'analyse : la dimension affective, la dimension historique, la dimension politico-stratégique et la dimension pratique.

La dimension affective.

Dans la première dimension, l'auteur met en avant la fonction de contrôle et de sélection et sa charge affective. Les mots "évaluation", "évaluateur", "évalué" évoquent spontanément d'autres termes, associés à des expériences affectives le plus souvent négatives.

La dimension historique.

L'évolution historique du concept d'évaluation et des pratiques (modèles dominants, normes, ...) s'explique pour Sensi par la conjonction de trois facteurs et de trois catégories d'acteurs : la pression du public, la rationalisation des dépenses publiques par les politiques et la professionnalisation de l'évaluation.

La dimension politico-stratégique.

Sensi justifie cette dimension par l'analyse stratégique du pouvoir dans les organisations, définie par Crozier et Friedberg. Elle identifie de la sorte une série de questions sur les buts et les acteurs de l'évaluation : qui demande et pourquoi, qui participe, qui contrôle et comment.

L'évaluation est un jeu stratégique avec des relations complexes de dépendance / indépendance. Sensi catégorise les utilisateurs en trois groupes : les SUPRA (décideurs politiques et administratifs), les INTRA (gestionnaires et participants) et les EXTRA (grand public, communauté scientifique : les utilisateurs secondaires).

Elle identifie trois fonctions principales à l'évaluation : le contrôle et la justification liés à la dépense de fonds publics (*accountability*), l'aide à la décision (décision politique ou de gestion de programme) et la diffusion d'informations (le "*droit de savoir*" de MacDonald).

La dimension pratique.

La dernière dimension présentée par Sensi concerne les aspects pratiques de l'évaluation. Elle y aborde : les objets, les questions et critères, les moments, les méthodes, l'analyse, la communication.

Les tâches à accomplir par l'évaluateur sont structurées autour de quatre étapes : la négociation qui, dans l'évaluation répondante, est une démarche permanente ; les méthodes de recueil des données qui sont interactives (naturalistes) ou non (positivistes) ; l'analyse des résultats quantitatifs et / ou qualitatifs ; la communication des résultats, qui est un processus plus complexe que la simple présentation d'un rapport final.

La démarche de Sensi est très détaillée. Tout en conservant une approche pratique assez classique, elle introduit une dimension stratégique essentielle, comme l'avaient souligné les études précédentes.

b) Méthode.

Sensi utilise plusieurs méthodes d'investigation. Elle analyse de manière quantitative et qualitative le contenu de 123 documents publiés par le Journal Officiel de la Commission européenne, afin d'identifier le discours au sujet de l'évaluation, son évolution sur 15 ans et les différences éventuelles entre les directions administratives.

Elle envoie aux unités administratives concernées par l'évaluation un questionnaire auto-administré composé de questions fermées portant sur les dimensions de l'évaluation.

Elle rencontre individuellement, pour un entretien semi-structuré approfondi, les 16 fonctionnaires ayant répondu au questionnaire.

c) Résultats.

A partir des différentes informations recueillies dans les textes, les questionnaires et les entretiens, Sensi propose une analyse historique, stratégique et pratique des évaluations menées à la Commission européenne.

Le rôle stratégique de l'évaluation, la partie la plus originale, est surtout mis en avant. L'évaluation prend une place qui croît en même temps que croît l'Europe. Elle suit l'évolution du nombre de programmes et l'évolution du rôle de la Commission face au conseil d'abord et au Parlement ensuite. L'évaluation est devenue une norme : elle est un enjeu pour la Commission et au sein de celle-ci, entre les Directions. L'évaluation économique est le modèle dominant à cause de son rôle de justification. Cependant, la diffusion des résultats est très limitée.

En même temps, cependant, se développe, dans le discours des fonctionnaires, une vision interne plus sensible à l'évaluation comme aide à la décision.

d) Conclusions.

Malgré l'introduction d'une dimension historique et stratégique, le travail de Sensi reste très descriptif. Les quelques informations récoltées sur les enjeux politiques et administratifs auraient bénéficié d'une discussion comparative avec les pratiques développées dans le contexte nord-américain.

En effet, Sensi étudie des pratiques d'évaluation à un niveau supranational où les caractéristiques socio-politiques des programmes jouent un rôle majeur. L'obligation d'évaluer s'inscrit dans un contexte de forte pression politique et administrative.

On peut faire le rapprochement avec le contexte des Etats-Unis, où les évaluations accompagnent certains programmes fédéraux. L'analyse aurait bénéficié d'une discussion de la littérature sur l'analyse de l'utilisation de ces évaluations, notamment de toute la réflexion sur le peu d'utilisation directe de l'évaluation dans un tel contexte.

Bien que ce travail fouillé apporte des suggestions sur un grand nombre de variables descriptives, le contexte de notre propre étude, les évaluations de programmes "de terrain", implique d'intégrer et de décrire une variété plus grande de formes d'évaluation.

2. Une enquête sur les pratiques en regard des "Standards".

Straw et Cook (1990) proposent de donner pour mission à la méta-évaluation l'étude non seulement de la méthodologie des évaluations, mais aussi de l'organisation, de la gestion et des effets de celles-ci. Ils citent comme exemple une étude de Marsch, Newman et Bover en 1981.

Marsch et al. (1981) ont utilisé les "Standards" proposés par le *Joint Committee on Standards for Educational Evaluation* pour étudier la pratique de l'évaluation en éducation en Californie. Dans cette étude, les standards sont perçus par les praticiens comme un idéal. La pratique courante est jugée de qualité moyenne en regard de ces critères. Les pratiques de basse qualité sont attribuées aux restrictions de temps, au politique ou à la bureaucratie et à l'incompétence de l'évaluateur.

Cette étude, plus proche de la pratique réelle, prend comme référence des normes professionnelles : l'utilité, la praticabilité, l'honnêteté, l'exactitude. En utilisant cette démarche, les pratiques jugées mauvaises ne sont pas définies par rapport à la pertinence, à l'efficacité réelle, mais par rapport à un idéal méthodologique. Les raisons invoquées sont dès lors exogènes (le temps, la politique, la compétence). Il n'y a pas d'explication sur la nature des évaluations entreprises ni sur les raisons de leur apparition.

3. Une enquête sur les pratiques des évaluateurs américains.

**Les pratiques d'évaluation de programmes d'éducation pour la santé
Chapitre 3 : recherches sur l'utilisation et la pratique de l'évaluation**

Shadish et Epstein (1987) remarquent l'abondance de conseils théoriques souvent contradictoires dans le domaine de l'évaluation. Ils se posent dès lors la question des pratiques réelles des évaluateurs. D'autant plus que les études disponibles restent centrées sur la description des méthodes utilisées sans tenter d'identifier ce que font les évaluateurs pour concevoir et réaliser leur travail.

Ces deux auteurs choisissent de définir les pratiques d'évaluations à partir d'une vaste enquête auprès d'un échantillon de 604 personnes parmi les 3.000 membres de deux associations professionnelles d'évaluateurs.

Leur questionnaire s'inspire de sept auteurs théoriques (Campbell, Cronbach, Rossi, Scriven, Stake, Weiss, Wholey) ; il comprend 74 questions portant sur un ensemble d'activités très large. Il inclut des informations descriptives sur la formation, l'identité démographique et professionnelle et sur la spécialisation en évaluation. Des questions sont également posées sur leurs expériences récentes, leurs évolutions, les contraintes rencontrées ainsi que sur l'influence des théories.

a) Résultats.

Trois cent cinquante personnes ont répondu. La comparaison du profil professionnel des répondants et des non répondants sur base du fichier des associations ne montre pas de différences significatives excepté une sous-représentation, parmi les répondants, des personnes travaillant en milieu universitaire.

Profil professionnel

- 70 % ont un doctorat ; 91 % l'ont obtenu il y a moins de 20 ans ;
- 60 % sont des pédagogues et des psychologues ; 8 % ont une formation spécifique comme évaluateurs ;
- 76 % travaillent comme évaluateurs ; plus d'un tiers supervisent, enseignent ou rédigent en matière d'évaluation ;
- 31 % font de l'évaluation leur première profession ;
- 41 % travaillent dans des agences publiques et 18 % pour le privé ;
- la majorité se sont formés par l'expérience directe et la lecture.

Evolution des pratiques

La moitié des répondants déclarent avoir profondément changé leur pratique depuis leur début professionnel en cette matière. L'origine de ces changements est d'ordre intellectuel : ils sont dus à l'apparition de nouvelles idées.

Contraintes

Les limites financières sont reconnues comme une contrainte majeure par près de 40 % des répondants. Cette contrainte pèse surtout sur la récolte des résultats (*outcomes*), puis sur le choix des méthodes.

Les commanditaires et les superviseurs influencent peu la manière de conduire l'évaluation. Cette influence est la plus faible dans le choix des groupes à interroger et dans le type de questions à explorer.

Influences théoriques

Un seul ouvrage est connu et influent. C'est celui de Campbell et Stanley sur les devis expérimentaux et quasi-expérimentaux. Deux autres ouvrages, celui de Rossi et Freeman et celui de Weiss, sont connus. Les autres ouvrages proposés (Scriven, Stake, Cronbach, Wholey) sont inconnus des répondants.

Les idées les plus influentes concernent les devis d'évaluation (Campbell), la distinction entre évaluation formative et sommative (Scriven), le rôle de l'évaluation dans l'amélioration des programmes (Cronbach), l'évaluation comme source de jugement sur la valeur d'un programme (Scriven).

Le mode d'influence privilégié de ces idées est leur convergence avec les valeurs et la personnalité de l'évaluateur, puis leur praticabilité et la concordance avec les besoins du travail de l'évaluateur. La plus grande validité d'une théorie par rapport à une autre n'est pas une raison de choix.

Pratiques d'évaluation

Nonante-deux pour cent des répondants ont pratiqué l'évaluation ; 73 % en ont réalisé entre 3 et 50.

Les 74 questions portant sur l'évaluation complète la plus récente se distribuent en 8 catégories. Quatre-vingt-neuf pour cent estiment que cette dernière évaluation est représentative de leur pratique et 60 % pensent qu'elle est aussi représentative de la meilleure façon de faire.

L'objectif.

L'objectif principal est la mesure des effets. Vient ensuite l'amélioration du programme et l'aide à la décision. Les objectifs plus scientifiques viennent pour terminer.

La décision de la réaliser.

Cette décision est surtout basée sur l'intérêt de démontrer le rôle de l'évaluation pour changer le programme et la curiosité scientifique. L'intérêt financier vient seulement ensuite.

Le rôle de l'évaluateur.

L'évaluateur se perçoit d'abord comme un expert méthodologique, un éducateur vis-à-vis de ses interlocuteurs. Il se considère ensuite comme un agent de changement. La position de soutien, d'aide vis-à-vis du gestionnaire du programme ou des acteurs vient en dernier lieu.

L'origine des questions traitées.

Ce sont les besoins en informations des commanditaires de l'évaluation qui sont à l'origine des questions traitées. La littérature sur l'évaluation et l'expérience viennent ensuite. Les besoins du gestionnaire, les décisions à prendre, les besoins des bénéficiaires suivent.

Le type d'informations recueillies.

L'implantation du programme, ses modalités d'application et de diffusion, apparaissent en première position, suivies par les bénéfices apportés par celui-ci.

L'origine des critères d'efficacité du programme.

Les buts du programme sont la première source de contrôle de l'efficacité. La deuxième source d'influence du choix des variables est la littérature, puis viennent les options des gestionnaires et enfin, en dernière position, les critères des commanditaires.

Les méthodes utilisées.

La première source d'information est fournie par le programme lui-même, par les données qu'il récolte et par l'observation sur le terrain. L'enquête auprès des bénéficiaires vient ensuite. Les méthodes quantitatives sont peu utilisées, ce sont par ordre décroissant : la mesure des besoins, les approches quasi-expérimentales, les tests de rendement et les expériences randomisées. La méta-analyse ferme la marche.

Les moyens de faciliter l'utilisation.

Le rapport écrit est la forme privilégiée de diffusion. Assez loin derrière, viennent la traduction des résultats en recommandations et les exposés oraux. Le contact plus direct ou le contact en cours d'évaluation sont moins utilisés. La publication dans des revues et la diffusion dans les médias sont peu fréquents.

Une analyse factorielle a permis de dégager quatre types de pratiques indépendantes. Ces quatre types expliquent 45 % de la variance

Deux variables sont présentes dans tous les facteurs. La première concerne le type d'informations à recueillir. Ce qui différencie les évaluateurs est moins ce qu'ils évaluent que la manière dont ils le font. La seconde décrit leur rôle. Tous se perçoivent comme des experts plutôt que comme des conseillers.

Le premier facteur regroupe les pratiques qualifiées d'académiques (*academic evaluation*) par les deux auteurs. L'intérêt pour l'évaluation est scientifique, l'objectif est de résoudre des problèmes sociaux et de faire progresser les connaissances. Les questions traitées et les critères d'efficacité sont issus de la littérature et des objectifs du programme. Les méthodes sont traditionnelles, quantitatives et méta-évaluatives. La diffusion est orientée vers la publication scientifique.

Le second facteur est appelé pratique de service (*stakeholder service evaluation*). L'évaluateur est rémunéré par un client. L'évaluation doit répondre à ses questions et fournir une information utile pour améliorer le programme. Les critères d'efficacité sont définis avec les différents acteurs. Le rôle de l'évaluateur est d'être un agent de changement et un partenaire de l'équipe du programme. Il utilise des méthodes comme l'observation, les entretiens, les études de cas. Il diffuse l'information de manière intensive et continue.

Le troisième facteur est une pratique d'aide à la décision (*decision-driven evaluation*). L'évaluateur se considère comme un consultant et pas comme un partenaire. Il envisage les bénéfices potentiels et les changements à la lumière des coûts. Il oriente son travail en fonction des décisions à prendre et des législations existantes. Il n'a pas de préférence de méthodes.

Le quatrième facteur est lié à l'évaluation des résultats, l'évaluation d'efficacité (*outcome evaluation*). Le but de l'évaluation est de juger la valeur et l'efficacité d'un programme. Le rôle principal est celui d'un expert, d'un éducateur et d'un juge. L'évaluateur utilise les méthodes de recherche et la récolte continue de données (monitoring). Il diffuse les résultats sous forme de rapports écrits et oraux. Il formule des recommandations.

Un regroupement avec un nombre différent de facteurs ne modifie pas la distribution des variables. La présence de trois facteurs voit le regroupement des facteurs 2 et 3 (pratique de service et pratique de décision) en un seul facteur à orientation de service. Le regroupement en deux facteurs voit s'opposer un facteur service et un facteur académique qui réunit la pratique académique et la pratique d'efficacité.

Selon les auteurs, ces facteurs n'expliquent cependant qu'une partie de la variance. Ils simplifient et schématisent les pratiques d'évaluation et ne prédisent pas le comportement d'un évaluateur en particulier, qui fait intervenir dans des proportions variables plusieurs de ces pratiques. Les facteurs ne permettent pas non plus d'identifier les pratiques les plus fréquentes. En effet, chacune des pratiques intervient partiellement dans l'explication de la variance de tous les items.

Analyse prédictive

Shadish et Epstein ont aussi réalisé une analyse prédictive des quatre schémas de pratiques en étudiant l'influence de la formation, de la situation professionnelle et des référents théoriques.

- Sur le plan de la formation, ceux qui ont un doctorat s'orientent vers une évaluation de type "académique" ou "efficacité". Ce sont aussi ceux qui s'identifient d'abord comme psychologues, économistes, politologues.
- Ceux qui s'identifient d'abord comme des évaluateurs utilisent l'approche de "service" ou de "décision". Il ont moins souvent un doctorat.
- Ceux qui se décrivent comme des pédagogues, des sociologues, des anthropologues sont répartis entre les différentes pratiques.

La situation professionnelle permet de prédire différentes pratiques. Les évaluateurs travaillant dans des agences publiques pour des états ou au niveau fédéral ont une pratique de "décision". Les évaluateurs travaillant pour des firmes privées spécialisées ou dans les universités ont une pratique "académique" et une pratique "efficacité". Les évaluateurs travaillant comme évaluateur interne au niveau local ou dans des firmes privées non spécialisées ont plutôt une pratique de "service".

La prédiction du type de pratique par l'influence de tel ou tel auteur n'est pas très significative. La pratique "académique" semble être influencée par Cronbach et la pratique de service par Stake, ce qui n'exclut pas l'influence d'autres auteurs. Les deux autres pratiques n'ont pas de liens significatifs avec un auteur.

b) Synthèse.

Les auteurs interprètent l'existence de différentes pratiques par des raisons dues à la situation professionnelle et à la formation. Mais la raison principale est pour eux la diversité même des situations d'évaluation qui nécessitent des réponses variées. La diversité est dès lors indispensable et bénéfique. L'évaluateur, à terme, devra pouvoir construire une évaluation adaptée (*tailored*) en choisissant la pratique la plus pertinente sur base des spécificités de la situation.

Ils nuancent cependant leurs propos optimistes en remarquant le faible niveau de familiarité avec les théories. Il y a là, selon eux, un danger important pour le statut professionnel de l'évaluation, que la raison en soit le faible niveau de lecture des évaluateurs ou l'incapacité des théories à résoudre les problèmes de la pratique. L'évaluation doit tendre à développer un corpus unique de connaissances applicables à tous les problèmes pratiques. Les évaluateurs doivent maîtriser ce corpus.

Ils concluent enfin en soulevant le risque d'un clivage entre une vision académique et une vision pratique de l'évaluation conduisant à une séparation professionnelle.

c) Conclusions.

Le travail de Shadish et Epstein est exemplaire. Ils utilisent une démarche très systématique dans la méthodologie d'enquête et l'analyse quantitative.

En s'intéressant aux pratiques réelles des évaluateurs, ils innoverent et apportent une image intéressante de la profession d'évaluateur aux Etats-Unis.

Les variables permettent d'approcher la pratique de l'évaluation du point de vue de celui qui la réalise : ses objectifs; les raisons de son intérêt; le rôle assumé; le choix des questions à traiter, des informations à recueillir, des critères d'efficacité; la méthode utilisée et le mode de diffusion préféré. En se focalisant sur la dernière évaluation réalisée, ils assurent une meilleure pertinence des réponses par rapport à la pratique réelle.

L'image obtenue est effectivement très "pratique", significative d'une profession en voie de développement. Les évaluateurs sont rarement formés de manière spécifique à leur métier. Ils sont un tiers à le faire comme première profession. La formation se fait sur le terrain et au moyen de lectures axées sur les comptes rendus d'expériences plus que sur des écrits théoriques. Ces théories, ils les utilisent plus en fonction de leurs besoins et leurs intérêts qu'en vertu de leur valeur propre. Ces évaluateurs sont jaloux de leur indépendance et de leur rôle d'expert.

Mais l'interprétation finale de ces deux auteurs est aussi exemplaire du postulat académique initial de ceux-ci. La théorie n'est pas seulement un point de départ permettant de proposer des variables descriptives des pratiques. Elle est aussi le point d'arrivée des conclusions. La théorie est indispensable à la pratique. Il faut aboutir à un corpus de connaissances unique que tout évaluateur doit connaître et utiliser. Rien dans l'étude ne semble pourtant conduire à cette conclusion.

Il est significatif qu'une majorité des évaluateurs adoptent une approche pragmatique, orientée vers le service plutôt que vers la recherche, se basant sur l'expérience. Nous avons rencontré ce phénomène dans notre analyse des modèles théoriques ou dans les résultats des recherches sur l'utilisation. Cette vaste étude sur les pratiques professionnelles le confirme. Elle nous renforce dans notre projet d'analyser les pratiques réelles, mais en travaillant dans un contexte différent, au cœur des programmes, et sans lien systématique avec la profession d'évaluateur externe.

Nous retenons l'interrogation initiale : "Que font les évaluateurs ?". Mais nous tentons de donner une définition plus large de l'évaluation et de l'évaluateur.

Nous retenons aussi le rôle marquant de la formation et de la situation professionnelle comme une clé d'interprétation des différentes pratiques rencontrées par Shadish et Epstein. La prédominance de certaines professions et l'absence d'autres (les médecins, par exemple) dans les affiliés aux deux associations étudiées par Shadish et Epstein a donc certainement un impact sur le type de pratiques qu'ils ont identifiées.

Les résultats de Shadish et Epstein confortent l'idée du rôle joué par la position interne ou externe vis-à-vis du déroulement d'un programme. Les employés des agences fédérales, des universités ou des firmes spécialisées en évaluation ont cette position externe et développent des pratiques différentes de leurs collègues plus proches de l'action. La complexité et la durée du programme conduit à un morcellement des tâches et à une spécialisation, qui implique elle-même de recourir à certaines pratiques externes. Cette spécialisation va de pair avec une plus grande fréquence d'énonciation, nécessaire pour justifier, rendre compte du travail spécifique réalisé, que ce soit face au gouvernement, face à la communauté scientifique ou face aux dirigeants de l'entreprise d'évaluation. La formalisation accompagne également ce phénomène de distanciation et de professionnalisation. Mais ces trois mécanismes - énonciation, externalisation et formalisation - restent distincts. Ils peuvent aller dans le même sens ou diverger en fonction des circonstances et des besoins de l'évaluateur.

4. Conclusions.

Les trois études que nous avons présentées permettent de situer notre propre recherche.

a) L'objet d'étude.

Pour tous, cet objet est l'évaluation, mais définie de manières différentes.

Sensi définit des descripteurs pour analyser les pratiques et les stratégies des fonctionnaires chargés d'évaluation dans les services spécialisés de la Commission européenne.

Pour Marsh, Newman et Bover, ce sont les "Standards" professionnels de l'évaluation qui permettent d'analyser les pratiques d'évaluation financées par l'Etat de Californie dans le domaine de l'éducation et menées par des professionnels.

Pour Shadish et Epstein, les pratiques d'évaluation des membres de deux associations professionnelles (*Evaluation Network et Evaluation Research Society*) sont confrontées aux propositions de sept auteurs sur la manière de mener une évaluation.

Quant à nous, nous avons appliqué un modèle global descriptif de la nature de l'évaluation aux projets et rapports produits par les gestionnaires de programme, nous éloignant ainsi volontairement de la pratique des professionnels.

b) La méthodologie.

Sensi fait une analyse de contenu qualitative et quantitative de textes officiels qui parlent d'évaluation, mais qui n'en sont pas. Elle interroge des évaluateurs professionnels par questionnaires fermés et entretiens semi-dirigés.

Marsh, Newman et Bover font une enquête par questionnaires auprès des responsables d'évaluation.

Shadish et Epstein font également une enquête par questionnaires auprès d'un échantillon de membres de deux associations professionnelles.

Nous faisons une analyse quantitative et qualitative des documents liés au programme d'action : le projet et le rapport, que nous complétons par une interview des gestionnaires de ces mêmes programmes.

CHAPITRE 4

L'EVALUATION DE L'EFFICACITE ET DE LA QUALITE EN SANTE PUBLIQUE ET EN PROMOTION DE LA SANTE

I.	INTRODUCTION: L'ANALYSE DE QUELQUES CAS.....	148
II.	L'EFFICACITE ET LA QUALITE EN SANTE PUBLIQUE	150
II.A.	LE TRANSFERT DE LA CONNAISSANCE VERS L'ACTION	151
II.B.	LA MESURE DE L'EFFICACITE EN SANTE PUBLIQUE ET EN SANTE AU TRAVAIL	157
II.C.	L'ACCREDITATION ET L'ASSURANCE DE QUALITE	161
II.D.	SYNTHESE SUR L'APPROCHE DE L'EFFICACITE ET DE LA QUALITE EN SANTE PUBLIQUE	162
III.	L'EFFICACITE ET LA QUALITE DE LA PROMOTION DE LA SANTE ET DE L'EDUCATION POUR LA SANTE	164
III.A.	POURQUOI POSER LA QUESTION DE L'EFFICACITE ET DE LA QUALITE ?	165
III.B.	COMMENT RESOUDRE LA QUESTION DE L'EFFICACITE ?	168
III.C.	COMMENT RESOUDRE LA QUESTION DE LA QUALITE ?	184
IV.	CONCLUSIONS : LA FECONDITE DU MODELE	194

I. INTRODUCTION : L'ANALYSE DE QUELQUES

CAS

Nous postulons l'unité de l'acte évaluatif. Il ne devrait donc pas y avoir de différences dans les pratiques d'évaluation de programmes des secteurs social, éducatif, sanitaire.

Une première analyse intuitive semble le confirmer. Dès le début du siècle, différentes formes d'évaluation (gestion, expertise, recherche évaluative, inspection) se retrouvent aussi bien dans le secteur de l'éducation que dans celui de la santé ou dans celui du social. Ainsi, en 1910, le rapport de Flexner, spécialiste de l'éducation, influence profondément le système de l'éducation médicale en Amérique du Nord (Floden, 1983). C'est un exemple de l'approche par expertise.

Certains des auteurs que nous avons cités appartiennent à des disciplines particulières. Leurs écrits concernent et intéressent néanmoins plusieurs secteurs.

- Suchman (1967), professeur de sociologie, qui travailla surtout dans le domaine de la santé publique, influença profondément l'évaluation en éducation.
- Leviton et Hughes (1981), qui ont fait la première revue de littérature sur l'utilisation de l'évaluation, sont des spécialistes des politiques de santé.
- A l'inverse, les ouvrages de Campbell et Stanley (1966), puis de Cook et Campbell (1979) sur les devis quasi-expérimentaux, conçus pour l'expérimentation dans le domaine social, restent une référence incontournable pour le chercheur en santé publique.

Des ouvrages de référence dans le domaine de l'évaluation en éducation pour la santé et en promotion de la santé (Windsor et al., 1984, Green et Lewis, 1986, Tones, Tilford et Robinson, 1990) citent abondamment les auteurs classiques de l'évaluation de programmes en éducation.

Battista et Tremblay (1985) notent que l'évaluation, issue d'une gamme étendue de disciplines, devient une profession particulière s'appliquant à une vaste gamme de programmes sociaux.

Monnier (1987, pp. 157 à 158), ingénieur et docteur en sociologie, suggère même que l'évaluateur ne soit pas un expert d'un secteur particulier de l'action politique (santé, éducation, habitat, ...), mais bien un généraliste de l'évaluation. Il plaide pour une communauté professionnelle de l'évaluation.

Ces quelques exemples ne suffisent cependant pas à établir si différents secteurs utilisent des "modèles" d'évaluation semblables. C'est pourquoi, avant de présenter la recherche sur les pratiques d'évaluation de programmes d'éducation pour la santé, il est utile de discuter de manière théorique la spécificité du champ de la santé. Le modèle général d'analyse des pratiques d'évaluation de programmes permet-il de rendre compte de diverses pratiques d'évaluation proposées dans la littérature de santé publique comme dans celle d'éducation pour la santé ?

Pour éviter une probable redondance dans la description de modèles théoriques, la grille APEP a été utilisée pour analyser deux problématiques qui reviennent couramment dans la littérature : l'efficacité et l'assurance de la qualité. Nous envisagerons d'abord ces deux questions dans le champ de la santé publique avant d'aborder celui de l'éducation pour la santé et de la promotion de la santé.

II. L'EFFICACITE ET LA QUALITE EN SANTE

PUBLIQUE

"Puisque la finalité du système de santé est de produire de la santé, l'approche la plus raisonnable n'est-elle pas d'essayer de mesurer en termes de santé - mortalité, morbidité, invalidité - les résultats des processus de soins ?" (Gremy et Goldberg, 1981)

La mesure des résultats de programmes de santé serait donc aussi "la mesure de la santé". Mais qu'est-ce que la santé et comment la mesurer ?

Le concept de santé et sa mesure n'ont cessé d'évoluer. C'était d'abord la mesure de l'absence de maladie, puis de l'état de santé, par des indicateurs biologiques et physiologiques. C'est, depuis peu, la mesure de *"l'état complet de bien-être physique, mental et social."* (OMS, 1948)

Cette définition, qui paraît encore à bien des égards utopique, introduit de nouveaux objets à mesurer. L'évolution de la technologie et de la médecine, la pression des organismes qui financent la santé et la demande croissante de la population pour une meilleure qualité de vie ont profondément modifié la perception collective de la santé et entraîné un développement d'autres modalités de mesure de la santé. La mesure de la santé se préoccupe de développer des indicateurs de plus en plus nombreux et complexes sur les capacités fonctionnelles, la perception de la santé et du bien-être, et même la qualité de la vie (travail, éducation, socialisation, etc.) (Pransky et Himmelstein, 1996).

Moatti (1992, pp. 127 à 129) synthétise cette évolution en proposant quelques définitions de l'efficacité pour le secteur de la santé dans une perspective économique.

"Ne produire que les services pour lesquels il existe une évidence clinique suffisante qu'ils sont efficaces, c'est-à-dire qu'ils contribuent effectivement à améliorer l'état de santé des patients." C'est la préoccupation du clinicien comme de l'épidémiologiste.

"Produire au moindre coût (compte tenu des contraintes techniques sur les facteurs de production) les services dont l'efficacité a été démontrée." C'est la préoccupation de l'économiste.

"Concentrer les ressources sur les services efficaces, produits au moindre coût, qui rapportent le plus en termes d'amélioration de la santé et du bien-être collectifs." C'est la préoccupation du "consommateur" de santé.

Efficacité technique, efficacité économique et efficacité psychosociale, telles sont les exigences actuelles vis-à-vis du secteur curatif. Mais ces logiques ne sont pas cumulatives, comme le voudrait une approche rationnelle telle que la propose Moatti. Elles sont souvent concurrentes.

II.A. LE TRANSFERT DE LA CONNAISSANCE VERS L'ACTION

1. La décision dans la pratique médicale et en santé publique.

a) La relation entre savoir éprouvé et pratique médicale.

Broclain (1994) note l'émergence d'un nouveau paradigme de la pratique médicale où l'expérience clinique individuelle devrait céder le pas à une pratique éprouvée par la recherche (*evidence based*), c'est-à-dire se basant sur des critères d'efficacité thérapeutique, d'efficacité économique et d'amélioration de la qualité de vie.

Ce nouveau paradigme est censé mieux guider la décision des praticiens. Mais la relation entre savoir éprouvé et pratique médicale est loin d'être "évidente". Quelles sont les raisons de ce hiatus ?

Tout d'abord, pour suivre la métaphore de Broclain (1994, p. 202), "*le savoir médical se présente comme un territoire composite où des îlots de terre ferme, parsemés de trous, côtoient de vastes zones marécageuses.*" Ensuite, face au cas individuel, la pratique médicale ne se réduit pas à un ensemble de faits aussi solidement établis qu'ils soient. Enfin, le comportement des professionnels de santé ne se modifie que lentement car il est basé autant sur la formation initiale, l'expérience individuelle et l'opinion d'experts que sur une analyse personnelle de la littérature scientifique ; d'autres considérations entrent également en jeu comme la structure hiérarchique de l'organisation hospitalière, la tradition, les positions éthiques, comme le montre Kutty dans son analyse de services de dialyse rénale (1973, cité par Crozier et Friedberg, 1977, pp. 272 à 277).

L'évaluation des interventions, aussi rigoureuse soit-elle, devient impraticable en situation réelle ou elle doit aller au-delà des preuves pour utiliser des présomptions, quand ce ne sont pas des convictions. Ces présomptions, ces convictions sont, en quelque sorte, les ponts jetés par les experts au-dessus des trous et des zones marécageuses que signalait Broclain.

Un des obstacles à une meilleure utilisation des connaissances de santé publique dans les processus décisionnels est la compréhension du rôle de l'expertise.

L'expert est au carrefour entre un savoir spécialisé et un problème à résoudre. Il reçoit un mandat en fonction d'une compétence reconnue. Par ce mandat, il est appelé non seulement à dire ce qu'il sait, mais aussi ce qu'il pense à partir de ce qu'il sait, à passer inévitablement du savoir à la conviction. Il est non seulement avocat de son savoir, mais aussi juge parce qu'il y a interprétation des données disponibles pour réduire l'incertitude, sans cependant parvenir à la supprimer (Roqueplo, 1991, cité par Dab, 1993, p. 205). L'expertise est un processus de jugement et non une activité scientifique. La formalisation de la démarche d'expertise, par exemple lors d'une conférence de consensus, n'ajoute rien à son caractère de jugement. D'autres formules d'expertise existent : désignation d'un consultant chargé d'une mission de synthèse, commission d'experts ou Conseil supérieur permanent ou encore expert désigné à l'opinion publique par les médias. Si la formule change, les difficultés subsistent.

La source de ces difficultés provient souvent de la confusion entre débats scientifiques et concertation des acteurs sociaux et économiques.

Toute étude, toute recommandation, qui entre dans le champ du jugement et de la décision déclenche des stratégies individuelles ou d'organisation : contestation de la validité, contre-étude, blocage, évitement, contrôle. La légitimité scientifique ne résout pas les conflits sociaux. Elle ne peut éviter la négociation (Dab, 1993).

Pour Monnier (1987), la contribution à la négociation devient le nouvel enjeu de l'évaluation.

b) Les relations entre l'information sanitaire et la décision en santé publique.

Une problématique constante est donc le transfert de la connaissance vers l'action. Dab (1993) présente une discussion très riche et très stimulante sur les relations entre l'information sanitaire et la décision en santé publique à partir d'exemples de surveillances épidémiologiques.

Pour Dab (1993), la santé publique se situe entre les savoir et les pratiques de santé. Elle est un champ d'action et de recherche. Elle doit aussi nécessairement se situer au niveau des valeurs et se faire "avocat de la santé". Dab pose ainsi d'emblée, pour la santé publique, le triple défi qui traverse notre étude : comment concilier connaissances, valeurs et décisions ? Mais c'est sur la relation entre le savoir et l'action qu'il centre d'abord son attention.

Comment expliquer, puis essayer de combler le fossé entre connaissance et action ?

On peut bien entendu se rejeter la responsabilité :

- les scientifiques peuvent se plaindre du manque de moyens, du manque de reconnaissance de la santé publique, du manque de formation en santé publique des médecins et des décideurs politiques et administratifs ;
- les décideurs peuvent accuser la santé publique de ne pas apporter les réponses utiles ou de le faire trop tard, dans un langage peu accessible, avec des résultats contradictoires, et assorties de longues mises en garde méthodologiques.

Pour tous, les études de santé publique sont longues et complexes à mettre en œuvre, les résultats souvent lointains. Mais ce débat est stérile et ne tient pas compte de la nature de la décision. Décider, c'est gérer l'incertitude. La connaissance produite par ces études ne tient pas compte du problème de la décision (Dab, 1993).

c) Des mécanismes pour assurer ce transfert entre connaissance et action.

Dab (1993) en applique quelques-uns au domaine de la santé publique. Il classe ces tentatives en deux grandes catégories.

La première met l'accent sur la connaissance pour faciliter son usage dans la décision. Il cite la démarche algorithmique, qui reformule les questions de recherche en fonction des préoccupations des décideurs ; la méta-analyse, qui cherche à synthétiser les informations ; et l'évaluation des risques, qui prend en compte l'incertitude dans l'information fournie.

La seconde formalise des règles d'utilisation des connaissances. L'analyse décisionnelle prépare la décision en étudiant des solutions alternatives. La conférence de consensus tente de dégager les implications pratiques des connaissances synthétisées. Seule cette dernière, selon Dab, ne part pas du présupposé de la possibilité de choisir une solution optimale. Cependant nous montrerons ci-dessous la différence entre le modèle théorique de la conférence de consensus et une application pratique (Broclain, 1994).

Dab conclut cette revue en affirmant qu'il n'y a pas de bonne ou de mauvaise méthode. L'essentiel est de ne pas masquer les enjeux sociaux et la transparence des choix.

Pour Dab (1993, p. 242), les modèles de décision s'appuyant sur l'idée de rationalité totale ne sont pas opérants en santé publique. Le lien entre connaissance et action ne peut être direct. Dès lors, Dab propose le concept de "rationalité accrue", c'est-à-dire l'explicitation des critères quantitatifs et qualitatifs utilisés dans les décisions (formalisation du jugement). Il ne nie pas la nécessité d'un lien entre connaissance et action, mais le but est moins d'accroître la rationalité des décisions que d'en comprendre et d'en faire partager le sens. Les décisions relèvent du champ du politique et pas de la science. Il n'y a pas de rationalité épidémiologique, comme il n'y a pas non plus de rationalité économique, qui puisse résoudre les problèmes de santé publique.

La solution réside alors dans le débat démocratique, dans l'explicitation des décisions et, dirions-nous, dans le processus qui y mène. La santé publique doit éclairer les enjeux, les conflits d'intérêts, les informations disponibles, les critères de jugement, les priorités, les choix possibles, les décisions et leurs conséquences.

C'est un choix éthique semblable qui se pose à l'éducateur pour la santé (Demarteau, 1991, Demarteau, 1993).

2. L'analyse d'une conférence de consensus.

a) La présentation et l'analyse de la conférence de consensus par Broclain (1994).

Les difficultés de coexistence entre la connaissance, le jugement et l'action sont illustrées très concrètement par l'analyse que réalise Broclain d'une conférence de consensus sur les œstrogènes et le risque de fracture.

Broclain commence par présenter le modèle de la conférence de consensus.

La procédure utilise trois modèles : celui de la réunion d'experts scientifiques, celui du procès judiciaire où un jury prononce un verdict et celui du forum public où les acteurs sociaux peuvent venir débattre leurs points de vue.

Selon les auteurs cités par Broclain (1994), comme Jacoby (1990) et l'ANDEM (Agence Nationale pour le Développement de l'Évaluation Médicale, 1990), la conférence de consensus a pour finalités :

- la synthèse des données scientifiques réalisée par un groupe non partisan (le jury) à partir de la discussion des experts (connaître) ;
- la résolution de conflits (juger) ;
- la modification des pratiques par la formulation de recommandations (décider).

Nous avons donc tous les ingrédients d'une évaluation "parfaite". Les conférences de consensus permettraient à la fois de connaître rationnellement, de juger impartialement et de décider réellement. Pourtant, ces conférences suscitent beaucoup de conflits, voire même de manipulations, comme le révèle Broclain.

Broclain analyse le déroulement de la conférence à partir des interventions des experts, des questions du jury et du public, de l'observation du fonctionnement du jury, des versions successives des conclusions de ce jury et à partir d'entretiens avec le comité d'organisation, les experts et les membres du jury.

Il présente les désaccords entre les experts et le jury sur le pouvoir prédictif de la densitométrie osseuse et sur l'efficacité anti-fracturaire des œstrogènes.

Ces désaccords portent sur la notion de "preuve suffisante" ; ils illustrent différents rôles de l'évaluation en cas de conflit entre des enjeux divergents.

La démonstration incontestable de la preuve d'une intervention nécessiterait la mise en œuvre d'un essai contrôlé randomisé débutant vers 45-50 ans et se poursuivant pendant une trentaine d'années. En l'absence d'une telle étude pour des raisons matérielles, techniques et éthiques, courantes en santé publique, les experts doivent extrapoler, dépasser la simple information scientifique pour se risquer à porter un jugement, à faire des recommandations, à proposer des décisions. Devant l'incertitude, les experts doivent quitter leur rôle scientifique et convaincre autant que prouver.

Dans la discussion, Broclain rappelle d'abord les enjeux médicaux, professionnels et sociaux de la médicalisation de la ménopause. Face à ces enjeux, le jury prend position comme défenseur de l'intérêt public en général et des femmes en particulier, choisissant une voie de prudence ("*d'abord ne pas nuire*") et s'interrogeant sur le coût social.

Le jury va donc prendre une position défensive et refuser de se laisser convaincre, en s'appuyant sur l'absence de preuve incontestable. Nous dirions que le jury, chargé d'émettre un jugement, va prendre la place des experts, chargés de dire la connaissance. Le jury utilise la connaissance pour contrer ce qu'il perçoit comme un jugement partisan des experts. C'est en quelque sorte l'histoire de l'arroseur arrosé.

Le rapport de force d'abord favorable au jury va s'effriter après un débat public houleux, lors duquel les experts rejettent les conclusions de celui-ci, mais surtout après la conférence elle-même, sous la pression de groupements professionnels médicaux menaçant d'organiser une nouvelle conférence de consensus.

En s'appuyant sur d'autres travaux sociologiques (Broclain, 1992, Wortman, Vinckur et Sechrest, 1988, cité par Broclain, 1994), Broclain (1994) conclut que la conférence de consensus ne permet pas de régler des controverses. Le jugement doit se baser sur des critères incertains. Si, pour se protéger, le jury agit autrement, il aboutit à une exacerbation du conflit. Prendre une décision en situation d'incertitude est caractéristique de la pratique réelle, que ce soit celle du médecin, de l'éducateur pour la santé ou du décideur politique.

Broclain propose des solutions pour faciliter le fonctionnement des conférences de consensus. D'une part, le jury n'a pas pour tâche de choisir des pratiques incontestables, mais de guider le praticien sur la solidité des pratiques. Il doit réaliser une appréciation interdisciplinaire, indépendante et transparente.

Pour Broclain, il faut mieux différencier les éléments prouvés et controversés. Il faut renforcer la reconnaissance de la compétence du jury. Il faut fournir l'explicitation des niveaux de preuves atteints en hiérarchisant la force des recommandations. Pour cela, il faut utiliser des grilles d'analyse critique de la qualité des publications. La lecture critique de l'information médicale est un point clé.

b) Les propositions de Broclain à la lueur de la grille d'analyse APEP.

En utilisant la grille d'analyse APEP, on remarque qu'une partie de ces propositions favorisent une plus grande explicitation des limites de l'information scientifique, elles restent donc dans le champ de la connaissance (une hiérarchie des preuves).

La conférence de consensus s'appuie sur des expertises et non sur un processus formel de récolte d'information. L'accent mis sur le traitement et la synthèse de l'information tend à renforcer le pôle connaissance, à passer d'un processus implicite à un processus explicite externe formalisé.

L'évolution du rôle du jury va dans le même sens, puisque Broclain propose une formalisation de la procédure de jugement en la rapprochant du pôle connaissance. La hiérarchie des preuves entraîne une formalisation du jugement par la construction des normes, de standards de la preuve scientifique. Il y a passage d'un jugement explicite externe informel à un jugement explicite externe formalisé.

La proposition d'un renforcement de la reconnaissance de la compétence du jury complète ce tableau. L'expertise fait partie de l'évaluation "jugement" ; cependant, la légitimité de l'expert ne viendrait pas de sa compétence à juger, ni de son expérience de praticien, mais de son savoir, expertise contre expertise.

Une autre partie des propositions de Broclain va aussi dans le sens d'une plus grande focalisation vers la décision (hiérarchie des recommandations). Il place la priorité sur l'évolution des pratiques. Toutefois la légitimité des recommandations reste issue de la science.

Les propositions de Broclain sont pertinentes dans la mesure où elles accentuent l'explicitation et la formalisation. Cependant le renforcement de l'analyse des preuves, de l'expertise, des recommandations ne modifie pas les enjeux ; cela peut faire croire que le rapport des forces peut se modifier si l'on renforce la légitimité du savoir. D'autres propositions auraient pu être formulées pour éviter cet écueil. Au lieu de construire les critères de choix uniquement autour de la preuve scientifique clinique, on pourrait introduire des critères économiques, sociaux et psychosociaux. On pourrait tenter de faire apparaître les enjeux professionnels, industriels, non pour les rejeter à priori, mais pour clarifier la négociation sociale et guider les décisions. C'est ce que propose Dab (1993).

La constitution du jury est à cet égard un point clé, qui reflète les divergences stratégiques. Broclain (1994) évoque diverses pratiques :

- Après avoir expérimenté, puis abandonné la formule d'un panel "équilibré" représentant les tendances existantes, les Américains prônent un jury constitué uniquement d'experts du domaine, mais considérés comme neutres.
- Les Britanniques ont opté pour un jury d'experts reconnus, mais externes au thème traité.
- La France a opté pour un modèle mixte, mais restant dans le secteur de la santé.
- Enfin, on pourrait imaginer un jury "populaire".

Sur le plan de la décision, on pourrait aussi proposer un resserrement et une plus grande formalisation du lien entre le jugement, les recommandations et l'expérience directe du praticien.

Grâce au travail minutieux de Broclain, l'analyse du cas de cette conférence de consensus illustre les relations complexes et variées entre la production d'un jugement de valeur, sa formulation et son utilisation dans le champ de la santé publique.

II.B. LA MESURE DE L'EFFICACITE EN SANTE PUBLIQUE ET EN SANTE AU TRAVAIL

L'analyse suivante envisage la question de la mesure de l'efficacité dans le domaine de la santé au travail (Pransky et Himmelstein, 1996). Quoique focalisée sur les objets de l'efficacité et sur les démarches d'évaluation de type formel, cette analyse montre le nécessaire élargissement de la réflexion sur les outils de mesure, proposant ainsi une transition vers la question de l'efficacité de la promotion de la santé et de l'éducation pour la santé.

Pransky et Himmelstein (1996) s'intéressent à l'évolution de la recherche sur les effets en matière de santé (*Outcome Research*) et à son impact sur la recherche dans le domaine de la santé au travail.

Les caractéristiques principales de cette évolution sont la prise en compte d'un nombre croissant d'effets ainsi que l'évolution des modèles et des outils pour les mesurer. Il ne s'agit plus de mesurer seulement la mortalité, la morbidité et des indicateurs biologiques. Il faut étudier aussi l'état de santé fonctionnel, les perceptions de la santé et la qualité de vie.

Pour cela, les mesures doivent se baser plus sur l'efficacité réelle que potentielle. En effet, l'une n'entraînerait pas l'autre. Il faut dès lors travailler avec d'autres outils et d'autres schémas de recherche que l'essai clinique randomisé : questionnaires, analyse de banques de données sociales et sanitaires, observations, prise en compte des intérêts de différents acteurs de la santé, ...

1. Le changement de paradigme du processus d'évaluation en santé publique.

Il y a là un changement de paradigme important. La démarche expérimentale ne serait plus le seul "standard" de la recherche en matière d'efficacité des interventions sanitaires. Pour comprendre cette évolution, il faut revenir sur les caractéristiques du modèle "expérimental" appliqué en santé publique.

"L'évaluation des interventions de santé publique est un processus scientifique qui permet d'apprécier dans quelle mesure celles-ci ont atteint les objectifs qui leur sont assignés." (Breart et Bouyer, 1991, p. 840)

"L'évaluation repose généralement sur la comparaison de deux groupes : un groupe de sujets bénéficiant de l'intervention (groupe A) et un groupe de sujets ne bénéficiant pas de cette intervention (groupe B). A partir de cette comparaison, on souhaite porter deux jugements : un jugement de signification (est-ce que les résultats dans le groupe A sont significativement meilleurs que dans le groupe B ?) et un jugement de causalité (est-ce que la différence de résultats constatée est expliquée par la différence entre les deux interventions ?)." (Breart et Bouyer, 1991, p. 841)

C'est le paradigme expérimental. L'expérimentation dans le champ social répond à cinq principes : relations causales, mesures comparatives, affectation aléatoire, représentativité et reproductibilité (Monnier, 1987).

L'application du paradigme expérimental au domaine social a montré ses limites, que nous avons évoquées au chapitre 2 dans la description des différents types d'évaluation. Ce modèle est-il plus valide en santé publique ?

La séquence rationnelle de la résolution d'un problème de santé publique est la suivante (Breart et Bouyer, 1991, pp. 846 à 847) :

- l'efficacité n'est pas prouvée. il faut des essais contrôlés randomisés sur l'efficacité de l'intervention ;
- l'efficacité est prouvée dans des conditions optimales. il faut une expérimentation en situation réelle avec ou sans randomisation ;
- l'efficacité est prouvée dans des conditions usuelles d'utilisation. il faut surveiller que l'on obtient les résultats escomptés durant la diffusion ;
- l'efficacité n'est pas prouvée, mais l'intervention est déjà généralisée. il faut des études d'observation.

L'évaluation des interventions en situation réelle est considérée comme une évaluation de routine. Elle est centrée sur la vérification de la conformité aux données acquises par l'évaluation de recherche, qui démontre l'utilité des interventions ou efficacité potentielle (Grémy et Goldberg, 1981).

2. L'évolution de la mesure des résultats dans le domaine de la santé au travail.

C'est cette séquence que remettent en cause les remarques de Pransky et Himmelstein (1996), qui postulent que le réel est aussi source de connaissance directe.

Pransky et Himmelstein proposent un examen de l'évolution de la mesure des résultats de santé dans le domaine de la santé au travail.

Pour la santé au travail, cette évolution implique un élargissement des effets à mesurer. Il faut tenir compte de l'état de santé fonctionnel et du retour au travail, mais aussi se préoccuper de la qualité de ce retour sur les plans fonctionnel et productif ainsi que du bien-être psychosocial.

La santé au travail a la particularité de poser le problème de la qualité de la vie pour une population jeune. Les indicateurs ne sont pas les actes de la vie quotidienne comme ceux mesurés auprès de populations plus âgées, mais bien le retour au travail productif. Le problème de la productivité prend le pas sur celui de la chronicité. Mais au-delà de cette productivité interviennent aussi des aspects moins strictement économiques comme le sentiment de sa propre valeur, les interactions sociales. La demande d'une meilleure qualité de vie avec ses aspects psychosociaux suppose un élargissement des concepts et des outils de l'évaluation de l'efficacité.

Pransky et Himmelstein proposent des pistes pour développer ces nouveaux outils de la mesure de la santé au travail :

- En santé au travail, la conceptualisation des effets à mesurer implique une investigation des champs de l'économie, de l'ergonomie, des sciences sociales, de la médecine du travail, des politiques de santé, des sciences administratives.
- Les définitions du travail, de la qualité de vie au travail doivent prendre en compte le point de vue des acteurs.
- Ce modèle multifactoriel nécessite de recueillir des données multiples (cliniques, sociales, économiques, psychologiques) auprès de services variés (assurances, sécurité sociale, hôpitaux).
- Des études longitudinales de longue durée sont nécessaires, mais se heurtent aux difficultés de ce type d'enquêtes : changement de travail, d'exposition aux risques, compliance à l'étude, ...
- Les patients doivent être représentatifs des situations de première ligne. Il ne faut donc pas créer des situations en vue d'études, mais étudier les situations réelles.
- Ces conditions supposent une collaboration avec les organisations de base comme les comités de sécurité, les médecins généralistes, les syndicats, ... la communauté.

Bref, il faut développer, selon eux, des stratégies variées d'évaluation. Cependant le but de ces auteurs reste de fournir aux différents acteurs (patrons, travailleurs, assureurs, médecins) les moyens de choix rationnels sur les stratégies alternatives en matière de prévention et de traitement. Ils restent résolument dans le champ de la connaissance ; ils confondent recherche et évaluation. Cette confusion est aussi présente dans une revue de littérature sur les recherches portant sur l'efficacité des interventions dans le domaine de la santé au travail et de la sécurité (Goldenhar et Schulte, 1994).

3. La qualité méthodologique des études d'intervention dans le domaine de la santé au travail.

Goldenhar et Schulte (1994) proposent une revue critique de la qualité méthodologique des "études d'intervention" (*intervention research*) produites de l'année 1988 à l'année 1993 dans le domaine de la santé au travail et de la sécurité. Leurs conclusions n'ont rien d'original et répètent le même constat que la plupart des travaux de ce type. Nous présenterons des conclusions semblables dans le domaine de l'éducation pour la santé. Nous en avons également trouvé dans d'autres champs de la prévention : toxicomanies, cancer, personnes âgées, etc.

Selon ces deux auteurs, les études manquent de base théorique, utilisent de trop petits échantillons. Les interventions sont insuffisamment décrites et trop peu intensives pour espérer entraîner un changement. Les schémas d'étude sont non expérimentaux ou quasi-expérimentaux et ne contrôlent pas assez les sources de biais.

Bien sûr ils ne suggèrent pas "*que les auteurs de ces études ne sont pas capables de conduire des recherches rigoureuses et valides*" ou "*qu'il est facile de mener une recherche sur une intervention avec une méthodologie solide.*" (Goldenhar et Schulte, 1994, p. 763) Leurs remarques ne seraient pas des critiques, mais des moyens de développer de nouvelles études qui permettent aux chercheurs et aux praticiens d'accumuler des preuves solides sur les meilleures pratiques.

Leurs suggestions sont multiples et pertinentes et témoignent d'un élargissement de la recherche, même si le modèle expérimental reste la référence. Les auteurs veulent à la fois la rigueur et la pertinence, l'objectivité et l'interaction avec les acteurs.

Malgré l'exactitude de ces analyses et la pertinence de ces suggestions, on ne peut manquer d'être interpellé par la récurrence de ce type de remarques, typiques de la pratique académique identifiée par Shadish et Epstein (1987).

A aucun moment, il n'y a de discussion sérieuse sur les raisons de la faiblesse chronique des études portant sur les interventions. Dès le départ, il y a confusion entre recherche et évaluation. Selon nous, le biais provient de la sélection des études reprises dans l'analyse.

La recherche sur les interventions est définie comme l'étude des interventions planifiées et mises en œuvre pour obtenir des résultats sur la santé. A aucun moment le terme "évaluation" n'est utilisé. Leur revue de littérature se base uniquement sur des mots clés évoquant des contenus : médecine du travail, prévention au travail, prévention des accidents, santé au travail, interventions sur la santé au travail et la sécurité, etc.

Pourquoi une étude d'efficacité d'un programme de santé au travail a-t-elle été publiée ? Par qui l'a-t-elle été ? Est-ce une recherche ou une évaluation ? Le but est-il de comprendre l'intervention, de porter un jugement de valeur sur celle-ci, de la guider ? Ces questions doivent au moins être posées avant d'appliquer les canons de la recherche aux études publiées. Le contexte, les stratégies influencent les études et par conséquent l'analyse secondaire de leurs méthodes et de leurs résultats.

C'est la principale ambiguïté de l'utilisation du terme "recherche évaluative" (*Evaluation Research*). Il masque un défaut de définition des objets à prendre en compte dans une revue de littérature et entraîne un biais de sélection qui invalide la plupart des remarques sur l'absence de qualité méthodologique des études sur l'efficacité des interventions. Ce manque de rigueur est patent dans un nombre important de publications utilisant indifféremment les termes : recherche, recherche évaluative, évaluation, études d'efficacité, recherches sur l'efficacité, recherches sur les interventions, etc.

Le modèle d'Analyse des Pratiques d'Évaluation de Programmes permet de montrer la similitude des problèmes soulevés par les interventions dans le domaine de la santé comme dans d'autres secteurs. Ils se posent au niveau du processus d'évaluation, du choix des objets à évaluer et des méthodes à utiliser. Ils se posent aussi au niveau de la formulation du jugement de valeur et de l'utilisation de celui-ci pour la décision. Ce modèle nous permet également d'analyser rapidement le cas de l'assurance de qualité dans le cadre de l'accréditation des hôpitaux, sur la base d'un article de Scrivens (1997) à ce propos.

II.C. L'ACCREDITATION ET L'ASSURANCE DE QUALITE

Scrivens (1997) analyse les systèmes d'accréditation des services de santé aux États-Unis, au Canada, en Australie et au Royaume-Uni.

L'accréditation d'un service de santé est un moyen de mesurer la qualité des processus d'organisation et des performances de ce service, en faisant contrôler par des inspecteurs le respect de normes fixées. Le plus souvent (États-Unis, Canada, Australie), ces normes sont fixées par les professionnels eux-mêmes et les services de santé paient un organisme d'accréditation indépendant.

L'accréditation est perçue comme un outil qui peut aider les financeurs (États, assureurs) à mesurer la qualité, la sécurité des services fournis.

Mais, selon Scrivens (1997), il faut distinguer deux concepts distincts bien que reliés : la qualité et la justification des moyens reçus (*accountability*). La qualité suppose une recherche constante de l'amélioration du service pour atteindre des normes de plus en plus élevées. La justification requiert la preuve publique d'un niveau adéquat de sécurité, d'efficacité et d'efficience. Nous dirions qu'il faut en somme distinguer "qualité" et "assurance de la qualité".

Scrivens (1997) montre qu'il existe différentes formes d'accréditation qui sont des écarts par rapport à un type idéal. Ce modèle idéal est caractérisé par une participation volontaire des services, des normes (standards) dont le respect est mesuré, ainsi que des évaluateurs externes et indépendants des organisations.

La grille APEP permet de définir ce modèle idéal comme un jugement explicite, externe et formalisé (**Jef**) dont le processus reste implicite de même que l'utilisation du jugement dans les décisions des services.

Pour Scrivens, il est clair que l'accréditation est un moyen parmi d'autres d'évaluer (*assessing*) les services de santé et les soins de santé. L'accréditation n'est pas une méthode scientifique, n'est pas un outil de mesure de la performance, mais est un système de gestion par consultance. Elle peut susciter des améliorations et agir comme un outil d'assurance de qualité, mais pas comme un instrument de mesure ou de surveillance de la qualité.

Au contraire, la pression politique pour mettre en place une "assurance de la qualité" conduit à choisir des indicateurs statistiques cliniques qui négligent les processus et donc à terme, l'amélioration de la qualité. L'accent est ainsi mis sur la surveillance de la sécurité et sur l'examen de l'utilisation des moyens financiers. Le système d'accréditation protège le public et rassure les financeurs. Il réduit par contre l'apprentissage professionnel qu'apportent des approches basées sur l'échange (*peer review*) ou l'amélioration interne (*self-improvement model*).

Ces différents modèles peuvent être décrits par notre grille d'analyse :

- le développement interne (*self-improvement*) est basé sur un jugement interne informel (**Jii**) ;
- l'examen par les pairs (*peer review*) est basé sur un jugement externe informel (**Jei**) ;
- les normes professionnelles recommandées (*guidelines*) mènent à un jugement interne formel (**Jif**) ;
- l'accréditation (*accreditation*) par un organisme indépendant est un jugement externe formel (**Jef**).

Cette différenciation permet d'examiner parallèlement le système de recueil d'information (processus) et l'impact sur la décision. C'est cet impact que discute Scrivens en indiquant que le système le plus externe et le plus formel conduit à une utilisation interne plus faible.

II.D. SYNTHÈSE SUR L'APPROCHE DE L'EFFICACITÉ ET DE LA QUALITÉ EN SANTÉ PUBLIQUE

Ce bref chapitre sur l'évaluation dans le domaine de la santé révèle les mêmes enjeux et les mêmes tensions entre les trois composantes processus – jugement – décision que dans d'autres secteurs sociaux. Le modèle d'Analyse des Pratiques d'Évaluation de Programmes permet de les mettre en évidence et de suggérer différents modes d'intervention.

Dab (1993) propose différentes stratégies de conciliation entre la connaissance et l'action : algorithme de décision, méta-analyse, évaluation des risques, analyse décisionnelle. Elles visent toutes à faciliter la communication avec les décideurs dans un cadre rationnel, qui néglige le rôle des valeurs en privilégiant celui de la connaissance. Au-delà de nouvelles méthodes qui permettent de renforcer le pôle de la connaissance est apparue l'importance d'explicitier les enjeux et les décisions ; est apparu également le rôle de l'évaluation dans les négociations.

Des démarches différentes étaient illustrées dans le chapitre 2, notamment par le modèle de Patton (1978) de l'évaluation orientée vers le client. Ce modèle part des questions des décideurs, tient compte des valeurs divergentes entre acteurs, puis cherche les connaissances utiles en fonction des décisions à prendre.

Le rôle de l'expert a été mis en évidence dans les stratégies de relation entre les savoirs et les pratiques, en particulier dans le cadre de la conférence de consensus. Il est difficile de situer la fonction exacte de cette démarche de consensus, entre la synthèse des connaissances, la résolution de conflits et l'évolution des pratiques.

Ces débats éclairent l'évolution de la recherche sur les effets en matière de santé au travail, thème pris comme exemple.

La connaissance des effets se doit d'intégrer une réflexion sur les pratiques réelles et sur les intérêts des différents acteurs en présence. Ce faisant, elle quitte le domaine du paradigme expérimental et celui de la rationalité, sans toujours le formaliser. En effet, le passage de la recherche à l'évaluation exige d'intégrer les dimensions de jugement et de décision, mais demande aussi une clarification sur la possibilité d'utiliser les connaissances acquises par l'évaluation pour enrichir le corpus scientifique.

L'intégration de nouveaux objets à mesurer et de nouvelles méthodologies, la prise en compte de nouvelles dimensions de l'évaluation demandent de nouveaux critères de qualité pour la sélection des études à intégrer dans les revues de littérature.

Enfin, les pratiques de l'accréditation, courantes dans le domaine des soins de santé, pourraient aussi bénéficier d'une clarification des différentes dimensions qu'elles recouvrent en vue d'en proposer une utilisation plus pertinente.

Ainsi, il semble que le modèle APEP permette de décrire, puis d'étudier la fréquence et les conditions d'apparition de multiples formes d'évaluation dans le champ de la santé. Cette catégorisation sera-t-elle aussi pertinente et utile pour approcher les pratiques d'évaluation en promotion de la santé et en éducation pour la santé ?

III. L'EFFICACITE ET LA QUALITE DE LA

PROMOTION DE LA SANTE ET DE L'EDUCATION

POUR LA SANTE

L'éducation pour la santé, la promotion de la santé sont-elles efficaces ? Donnent-elles l'une et l'autre des garanties de leur qualité ? Ces deux questions sont au centre de bien des débats.

La première est déjà ancienne. Modolo et Mamon (1991) font une revue des principaux thèmes développés au cours des 40 ans d'existence de l'Union Internationale de Promotion de la Santé et d'Education pour la Santé (UIPES) à travers les actes des conférences mondiales qui se sont tenues tous les trois ans. Parmi les thèmes principaux figurent l'évaluation et la recherche.

L'évaluation de l'efficacité a toujours été présente comme thème de débats au sein de l'Union (1953, 1956, 1962, 1969, 1976, 1979, 1982, 1985). La nécessité de la recherche fondamentale et appliquée est aussi une constante (1959, 1962, 1969, 1982, 1985) et l'existence d'un fossé entre la recherche et la pratique est dénoncée dès 1972.

Parmi les objets à mesurer, c'est la mesure des changements comportementaux qui apparaît la première (1962), viennent ensuite les indicateurs de l'état de santé et les coûts (1979, 1982, 1985).

Sur le plan méthodologique, l'accent est mis, à la même époque, sur l'établissement d'une ligne de base et la comparaison (1962). Dès 1969, on souligne l'importance de l'évaluation de l'efficacité à moyen terme pour déterminer le maintien du comportement. L'intérêt de l'échantillonnage aléatoire est surtout mis en avant en 1985.

Dans les années quatre-vingt, de nouvelles préoccupations voient le jour : comprendre comment l'éducation fonctionne, étudier le transfert des résultats de situations expérimentales à des programmes larges, mesurer des indicateurs de société, utiliser des méthodes qualitatives, se préoccuper des méthodes de diffusion des connaissances vers la pratique. L'évaluation de processus ne devient une préoccupation qu'en 1985.

Le lien entre l'évaluation et la planification est souligné depuis 1959, puis les propositions se développent : évaluation des besoins (1962), diagnostic éducationnel (1972).

Ce tour d'horizon de Modolo et Mamon (1991) est significatif de la permanence de la question de l'efficacité et de son ancrage dans la planification, mais aussi de son évolution. L'évaluation élargit l'éventail des indicateurs d'efficacité, se préoccupe du processus et s'intéresse aux méthodes qualitatives.

Le problème de l'efficacité prend encore plus d'ampleur en Europe à partir de 1989, à l'initiative des Pays-Bas qui vont organiser, à Rotterdam, la première conférence européenne sur l'efficacité (*First European Conference on Effectiveness of Health Education*). Une deuxième se tiendra en Grèce en 1992 et la troisième en Italie en 1996. C'est au cours de celle-ci que le thème de l'assurance de qualité apparaît comme une nouvelle priorité.

La question de l'assurance de qualité est donc beaucoup plus récente. Rootman (1997) indique que la première réunion sur l'assurance de qualité en Promotion de la Santé au Canada a lieu en 1996.

Les questions de l'efficacité et de la qualité seront abordées en deux temps sur la base d'une analyse de quelques articles représentatifs de différentes positions. Tout d'abord, quelles sont les raisons possibles de l'importance accordée à l'efficacité, puis à la qualité en promotion de la santé et en éducation pour la santé ? Ensuite, quelles réponses ont été apportées ?

III.A. POURQUOI POSER LA QUESTION DE L'EFFICACITE ET DE LA QUALITE ?

Pourquoi la question de l'efficacité est-elle posée depuis si longtemps en éducation pour la santé et en promotion de la santé sans trouver de réponse ? On peut tenter plusieurs hypothèses :

- la preuve de cette efficacité n'a toujours pas été apportée ;
- la question évolue et les méthodes pour apporter cette preuve sont encore à découvrir et à mettre en œuvre ;
- la question n'attend pas de réponse.

Absence de réponse, question et réponses en évolution, ou absence de question, les trois hypothèses contiennent peut-être une part d'exactitude ?

Il est difficile à l'éducation pour la santé et davantage encore à la promotion de la santé de montrer de manière scientifiquement incontestable, par la méthode expérimentale, leurs effets sur l'état de santé de la population.

Comme ce fut évoqué pour la mesure de la santé, la perception de la santé aujourd'hui n'est pas celle d'hier. La question de l'évaluation de l'efficacité évolue parallèlement à l'évolution des objectifs de l'éducation pour la santé et de la promotion de la santé et leurs moyens d'interventions.

Enfin, la question de l'évaluation de l'efficacité doit faire l'objet de clarification au regard des différentes composantes de l'évaluation. Derrière cette question, il y a le plus souvent celle du jugement de valeur sur le programme. Prouver l'efficacité demande un examen des enjeux et des acteurs (les valeurs), une réflexion sur les moyens de la preuve (le processus) et une interrogation sur l'utilisation des preuves (décision). Une preuve expérimentale de l'effet d'une action d'éducation pour la santé peut être triviale et ne pas faire progresser la connaissance des meilleures pratiques d'action. Elle peut ne répondre à aucun enjeu particulier et ne conduire à aucune utilisation. Dans ce cas, la forme de la preuve aurait plus d'importance que son contenu et sa portée.

Toutes ces pistes de réponses sont à explorer.

Néanmoins, les analyses proposées par de nombreux auteurs éclairent certains enjeux. Nous en présenterons trois. Les deux premières ressemblent à celles formulées dans la première partie de ce chapitre : elles concernent l'efficacité clinique et l'efficacité économique (Moatti, 1992). La troisième est plus spécifique : elle s'intéresse au statut de l'éducation pour la santé et de ceux qui en ont fait profession au sein du système de santé.

1. L'efficacité clinique.

L'efficacité clinique sert à sélectionner des activités de promotion de la santé et d'éducation pour la santé pour lesquelles il existe une évidence clinique qu'elles contribuent à améliorer la santé.

Kok (1989) affirme, par exemple, qu'il faut s'appuyer sur des bases épidémiologiques et psychologiques sûres sous peine d'être inefficace ou même contre-productif. La *Society of Health Education Officers* au Royaume-Uni (Macdonald et Adams, 1986, cité par Piette, 1990) propose, comme premier point d'un code éthique, la protection de la population vis-à-vis des interventions inutiles, contre-productives, invasives.

C'est donc une exigence professionnelle et éthique. La promotion de la santé ne doit pas nuire aux personnes et aux groupes auxquels elle s'adresse. Elle doit améliorer effectivement la santé. Cette exigence, apparemment simple, est en fait complexe à mettre en œuvre en promotion de la santé.

Pour la rencontrer, il faut résoudre le problème de la définition de la santé et de la définition de l'évidence clinique. Suffit-il, par exemple, de définir l'efficacité (l'évidence) comme l'atteinte des objectifs préétablis (Veen et al., 1994) ? Cette définition ne fait que reporter le problème vers la formulation des objectifs et des indicateurs, vers les définitions ou les "idéologies" de la promotion de la santé et de l'éducation pour la santé (Tones, 1986, 1992). Pour Green (1989), la première question à résoudre pour l'évaluation de l'efficacité est le choix des indicateurs et des critères ; c'est plus un problème administratif et politique que méthodologique.

2. L'efficacité économique et politique.

Ce sont les arguments le plus souvent utilisés par les auteurs (Tones, 1992, Veen et al., 1994, Macdonald, 1996, Ovretveit, 1996) pour justifier le développement de l'évaluation d'efficacité ou de l'évaluation de la qualité, qu'ils nomment plus souvent recherche évaluative et assurance de qualité.

La recherche sur l'efficacité est surtout nécessaire pour montrer aux décideurs que l'éducation pour la santé obtient des résultats (de Haes, 1991). Cette position est clairement exprimée par Tones (1992, p. 13) :

"If we are to prosper and survive we must demonstrate to the politicians who control our resources that we have been successful."

Cette obligation passe par des justifications économiques. La question économique dans le domaine de la santé est devenue une priorité politique. L'éducation pour la santé et la promotion de la santé se situent dans un contexte attentif aux coûts et dans un contexte de réduction des dépenses et de choix dans l'allocation des ressources. Cette pression économique se manifeste à plusieurs niveaux :

- il faut montrer tout d'abord que les services, les programmes sont produits au moindre coût. C'est le bon usage des moyens accordés (Deccache, 1997) ;
- il faut montrer aussi un rapport coût-efficacité favorable dans la compétition avec d'autres services de santé (Ovretveit, 1996).

Mais l'intention des responsables politiques est aussi de réduire les coûts de la santé par la prévention, la promotion de la santé et l'éducation pour la santé (Veen et al., 1994). A la question de l'efficacité de la promotion de la santé qui existe depuis longtemps, se superpose la question de son efficacité à réduire les dépenses de santé.

Toutefois, ce discours est-il crédible face à la réalité des sommes engagées dans le curatif, dans le préventif et, au sein du préventif, dans la promotion de la santé et l'éducation pour la santé ? Ces sommes sont dérisoires et les enjeux également. Les choix économiques et politiques se situent bien en amont de la promotion de la santé (Deccache, 1997).

Le discours sur l'efficacité est peut-être aussi à situer dans le contexte de la professionnalisation du secteur de la promotion de la santé.

3. L'enjeu professionnel.

La recherche sur l'efficacité et sur la qualité apparaît pour certains comme une mode. Elle est perçue comme un proposition de spécialistes universitaires inutilisable sur le terrain (Lafontaine, 1990) ou comme une demande des gestionnaires de la santé à situer dans le climat "libéral" actuel (Saan, 1997). L'obligation de prouver l'efficacité serait imposée de l'extérieur et d'autant plus facile à exiger qu'elle s'adresserait aux nouveaux venus de la promotion de la santé (Saan, 1997).

Cependant l'évaluation correspond aussi à une demande interne au secteur de la promotion de la santé et de l'éducation pour la santé. Baric (1980) remarque que la demande d'évaluation provient d'éducateurs pour la santé spécialisés travaillant dans des services spécialisés ; les uns et les autres doivent justifier leur existence et les moyens qui leur sont attribués.

En parlant de l'avenir de l'évaluation en promotion de la santé, Macdonald (1996) justifie son utilisation par la nécessité de se faire valoir par rapport au secteur curatif sur le plan de la rigueur des méthodes.

Ces indices mettent sur la voie du rôle probable de l'évaluation comme enjeu de la reconnaissance professionnelle de l'éducation pour la santé et de la promotion de la santé au sein du système de santé.

Piette (1990) met bien en évidence ce phénomène au niveau international et au sein de la Communauté française de Belgique. La reconnaissance professionnelle passe notamment par une formation spécifique et le statut de discipline scientifique. Ce fut le cas aux Etats-Unis, au cours des années soixante, sous l'impulsion de la *Society for Public Health Education*. Cette Société assure une défense professionnelle qui passe par le développement et la reconnaissance des compétences. Une dynamique de ce type est perceptible en Communauté française depuis le développement des formations universitaires au début des années quatre-vingt, bien que d'autres signes de la professionnalisation, comme la création d'une organisation professionnelle, ne soient pas encore apparus.

Nguyen-Nam (1990) précise la place du discours scientifique dans l'émergence de l'éducation pour la santé comme nouveau secteur socio-sanitaire au sein de la Communauté française. Cet auteur remarque la place que prennent les scientifiques dans une double légitimation. D'une part, la rigueur scientifique légitime les activités aux yeux des pouvoirs politiques. D'autre part, elle permet une légitimation comme discipline spécifique au sein du système de santé et de la communauté scientifique.

La valorisation de l'éducation pour la santé et des professionnels qui la pratiquent doit se faire par rapport à l'activité curative et au modèle socio-médical dominant. Elle passe par le canal de la compétence professionnelle basée sur une formation scientifique pour les acteurs comme pour les experts. Sans être un facteur décisif, ce contexte de professionnalisation influence certainement la permanence du discours sur l'évaluation de l'efficacité d'abord, sur l'assurance de qualité ensuite.

III.B. COMMENT RESOUDRE LA QUESTION DE L'EFFICACITE ?

Au début des années septante, Stufflebeam se penche sur l'état de santé de l'évaluation en éducation et la déclare bien malade (Stufflebeam et al., 1971). Dix ans plus tard, l'évaluation de l'efficacité en promotion de la santé et en éducation pour la santé fait également l'objet d'un diagnostic de carence.

En 1980, Davidson réalise une revue de littérature sur les programmes qui visent à modifier les styles de vie. Il remarque que les évaluations n'ont pas d'impact sur les programmes. Il en conclut que tenter de modifier les pratiques professionnelles par de l'information sur les résultats est naïf sur le plan psychologique et simpliste sur le plan des organisations (Davidson, 1980, cité par Bury, 1988).

Face à la récurrence de ce constat d'une absence de transfert entre l'évaluation et la pratique professionnelle (Ledwith, 1986, Kok, 1989, Veen et al., 1994), des auteurs ont tenté d'en diagnostiquer les causes et de proposer des remèdes.

Les causes identifiées et les remèdes proposés peuvent être regroupés en deux grandes catégories. La première est l'existence d'un "fossé" entre les évaluateurs d'une part, les décideurs et les praticiens d'autre part, entre le savoir et l'action. La seconde concerne la "spécificité" de la promotion de la santé et de l'éducation pour la santé dans le champ de la santé.

1. Réduire le "fossé".

Le peu d'utilisation de la recherche et de l'évaluation serait dû à un manque de communication et à un manque d'intérêt mutuel.

Dans le cadre d'un séminaire international organisé par l'OMS, Nutbeam (1987) utilise le schéma de la communication pour identifier 10 problèmes d'échanges entre producteurs et utilisateurs de résultats de recherche en promotion de la santé. L'année suivante, un autre séminaire international organisé par l'UCL (Deccache, 1988) réunit des chercheurs et des décideurs sur le thème de l'évaluation en éducation et établit une liste de recommandations. Piette (1990) ne voit pas seulement cet écart entre chercheurs et praticiens, mais entre praticiens "banalisés" et "spécialisés". Elle l'explique par une professionnalisation accrue, de nature scientifique, d'une partie des praticiens.

La promotion de la santé se trouve bien entre le savoir et l'action, dans la position décrite par Dab (1993) à propos de la santé publique.

Nutbeam (1987) propose de réduire le fossé en améliorant la communication entre les producteurs de recherches et les utilisateurs. Nous retrouvons les propositions déjà rencontrées et décrites dans les chapitres précédents telles que la revue de littérature, la méta-analyse, la conférence de consensus, l'analyse décisionnelle, etc. Ces initiatives visent à rendre les données plus accessibles, à les synthétiser, à mieux cibler les utilisateurs, à mieux diffuser les connaissances. Bref, il faut faire de la communication des résultats un objectif particulier.

On trouve en éducation pour la santé et en promotion de la santé de nombreuses revues de littérature sur des thèmes ou des publics particuliers, par exemple sur les personnes âgées (Rimer et al., 1986) mais aussi des méta-analyses. Iverson et Levy (1982) font des recommandations pour l'utilisation de la technique de la méta-analyse en éducation pour la santé.

Par contre, nous n'avons identifié dans la littérature aucune utilisation de la conférence de consensus en promotion de la santé et en éducation pour la santé. Une démarche proche a été expérimentée à l'Université de Liège pour permettre la discussion publique et contradictoire des résultats de l'évaluation d'un programme de prévention du tabagisme en milieu scolaire. (Monoyer, Miermans et Demarteau, 1994).

On trouve aussi en Europe plusieurs tentatives de revue exhaustive des résultats de l'évaluation de l'efficacité (Gatherer et al., 1979, Liedekerken et al., 1990, Veen et al., 1994). La première a été réalisée par le *Health Education Council* au Royaume-Uni et la seconde par le *Dutch Health Education Center* au Pays-Bas. La troisième, la plus ambitieuse, fut menée à l'initiative du Bureau régional européen de l'Union Internationale de Promotion de la Santé et d'Education pour la Santé (UIPES), coordonnée par le Centre Néerlandais de promotion de la santé et financée par l'Union européenne. Cette vaste revue de la littérature sur l'efficacité aboutit à une base de données et à une série de publications de synthèse centrées sur 16 thèmes : nutrition, sida, santé mentale, santé bucco-dentaire, drogues, tabac, accidents, exercice physique, cancers, maladies cardio-vasculaires, personnes âgées, patients à l'hôpital, patients en dehors de l'hôpital, pauvres et immigrés, milieu de travail, milieu scolaire.

Cet exemple, qui est analysé ci-dessous, montre les difficultés auxquelles se heurtent les tentatives de "réduire le fossé" par des revues de littérature sur l'efficacité. Il illustre aussi les nouvelles solutions qui tiennent compte des spécificités de la promotion de la santé et de l'éducation pour la santé. En effet, et c'est remarquable, la démarche fit l'objet d'une analyse interne et fut soumise à la critique d'un comité réunissant analystes et experts invités (Veen, 1995).

Pour sa part, Nutbeam (1987) va plus loin en demandant au chercheur de s'impliquer dans le débat public, de communiquer plus largement vers le public. Il propose aussi de créer un dialogue permanent entre décideurs et chercheurs par des contacts directs, réguliers, dès les premiers stades de la recherche et tout au long de celle-ci. Il rejoint là les propositions formulées par les recherches sur l'utilisation de l'évaluation (Cousins et Leithwood, 1986). Il propose aussi de répondre aux questions des décideurs, comme le fait l'évaluation orientée vers le client (Patton, 1978).

On retrouve, parmi d'autres propositions d'ordre méthodologique ou éthique émises par la conférence de Bruxelles (Deccache, 1988), des éléments convergents avec les recommandations formulées ci-dessus : susciter le dialogue, prévoir la diffusion dans le contrat d'évaluation, négocier le protocole d'évaluation avec le décideur, reconnaître la place de l'évaluation dans la démarche politique, assurer une communication précoce et active avec le décideur et vers la population.

Cependant, ces recommandations ne semblent pas avoir fait l'objet d'études particulières.

2. Un outil d'analyse des études d'efficacité de la promotion de la santé et de l'éducation pour la santé.

Le point de départ de la démarche de Veen et al. (1994) est le constat d'un manque d'informations sur l'efficacité, qui reprend les remarques habituelles :

- carence en études d'évaluation de bonne qualité ;
- manque de publications ;
- problème d'accessibilité (dispersion et problèmes de langue) ;
- manque de consensus sur les critères de mesure de l'efficacité en promotion de la santé et en éducation pour la santé.

Il faut donc récolter des informations sur les effets et sur la manière de les mesurer. Cela permettrait de mieux connaître l'efficacité pour définir les priorités, les méthodes d'intervention et d'évaluation.

- a) Le projet consiste à collecter des informations sur l'efficacité et sur la qualité de la recherche en évaluation et à la rendre plus accessible par une base de données et une série d'articles d'analyse.

Une grille d'analyse permet de décrire les méthodologies et les résultats des études d'évaluation publiées. Elle fut construite à partir de la littérature sur l'évaluation et la planification, particulièrement sur la base du modèle *Precede/Proceed* (Green et Kreuter, 1991).

La grille fut d'abord soumise aux futurs analystes, un groupe d'experts européens, puis remaniée selon leurs remarques. Elle comprend finalement 29 questions ouvertes et fermées réparties en 6 rubriques : description de l'intervention, description de l'évaluation, résultats de l'évaluation sommative, résultats de l'évaluation formative, recommandations, estimation de la qualité par l'expert.

Les analystes devaient appliquer la grille à 10 articles sélectionnés par eux sur base de critères tenant compte de la qualité de l'intervention et de la qualité de l'évaluation (groupe de contrôle, mesures pré et post).

Ils devaient ensuite rédiger une synthèse sur les résultats en matière d'efficacité et sur la qualité des études, puis faire des recommandations. Il leur était demandé aussi de commenter l'usage de l'instrument d'analyse à l'aide d'un questionnaire.

- b) Veen et al. analysent le résultat de la démarche en s'appuyant sur la manière de remplir les grilles d'analyse, sur le contenu de ces grilles et sur les commentaires des analystes.

La démarche d'analyse pose manifestement des problèmes, qui seront mieux identifiés l'année suivante, lors d'un séminaire entre analystes et experts.

Veen (1995) introduit le compte rendu de ce séminaire en réaffirmant l'utilité de la démarche pour prouver l'efficacité de la promotion de la santé et de l'éducation pour la santé auprès des décideurs politiques et pour guider les interventions. Mais dans une analyse théorique, il introduit de nouveaux concepts en remettant en cause le modèle dominant de l'évaluation d'efficacité (expérimental et quantitatif). Il met en évidence des difficultés pratiques dues à la complexité des situations "naturelles" rencontrées en promotion de la santé. Il propose de développer les méthodes qualitatives, l'approche "naturaliste", l'étude du processus.

Ainsi , la promotion de la santé fait sa révolution phénoménologique près de 25 ans après l'évaluation en éducation.

L'analyse plus détaillée de la démarche de cette revue de littérature met en évidence plusieurs difficultés (Veen, 1995).

La principale semble liée aux critères de sélection des publications analysées. Ces critères ont favorisé les études en langue anglaise, principalement américaines, qui utilisent une démarche expérimentale ou quasi-expérimentale quantitative. Ce type d'étude laisse peu de place à la description de l'intervention, considérée comme une boîte noire et, par conséquent, elle néglige l'évaluation du processus. Parmi les publications analysées, il n'y a pas d'intervention complexe faisant intervenir plusieurs stratégies, plusieurs publics, car les variables sont difficiles à isoler. On y trouve des évaluations de connaissances, d'attitudes, de comportements, mais rien sur les aspects sociaux, politiques ou environnementaux. On remarque également, probablement pour la même raison, l'absence d'études longitudinales et d'évaluation d'effets à long terme. Dans une recherche évaluative, ce qui guide le plus souvent le chercheur n'est pas l'évaluation de l'intervention, mais la réponse à ses questions et les contraintes matérielles et académiques immédiates dans lesquelles il se trouve. Le manque d'informations sur les effets absents ou négatifs, la petite taille des effectifs utilisés sont peut-être aussi des conséquences de ces contraintes.

De plus, non seulement les publications ne reflètent qu'un aspect de l'éducation pour la santé, mais la grille d'analyse ne paraît pas adéquate aux experts pour rendre compte des activités de promotion de la santé et des évaluations plus qualitatives (Veen, 1995). On ne peut dès lors que s'inquiéter du pouvoir normatif d'une telle grille, comme le dénoncent certains analystes (Veen et al., 1994), d'autant plus que Veen (1995) propose de l'utiliser pour concevoir, sélectionner et financer de nouveaux programmes.

En conclusion, le séminaire formule des recommandations intéressantes (Veen, 1995) pour le développement de l'évaluation :

- des recommandations techniques : clarification des concepts, formation des analystes, utilisation de deux analystes par publication, utilisation d'une hiérarchie des méthodologies de preuve, en ce comprises les méthodes qualitatives ;
- la modification des critères de sélection des publications pour inclure le non-expérimental et le qualitatif ;
- la stimulation des évaluateurs à l'utilisation de l'évaluation de processus et la stimulation des comités de lecture des journaux scientifiques à les accepter ;
- la collecte d'informations sur les interventions auprès des réalisateurs de projets et non plus seulement dans les publications.

3. Tenir compte de la spécificité.

De nombreux auteurs se sont attachés à expliquer les difficultés de l'évaluation de l'efficacité par la spécificité de la promotion de la santé et de l'éducation pour la santé dans le champ de la santé.

Ils font tout d'abord remarquer que la promotion de la santé et l'éducation pour la santé couvrent un domaine très large d'objectifs, de stratégies et de méthodes d'intervention. Au point qu'il semble impossible d'assurer une synthèse, un consensus entre les pratiques. C'est un des constats de la revue de la littérature sur l'efficacité menée par Veen et al. (1995).

Cette absence de consensus se reflète inévitablement dans la méthodologie de l'évaluation. Les indicateurs, les critères, les instruments de l'évaluation sont fonction de la définition du succès que l'on vise (Tones, 1992).

Le terme évaluation recouvre aussi des significations variées. Scientifiques, politiciens et praticiens posent différemment la question et ont différentes attentes, aussi valides les unes que les autres (Nutbeam, Smith et Catford, 1990).

Tones (1986) a proposé une classification des idéologies de l'éducation pour la santé : préventive, sociale, éducative et autonomisante. Il montre comment ces idéologies conduisent à choisir des indicateurs différents, qu'il appelle "intermédiaires" (Tones, 1992).

Pour structurer la présentation des difficultés de l'évaluation dans le champ de la promotion de la santé et de l'éducation pour la santé, nous aborderons successivement la question du choix des objets à évaluer et des contraintes imposées par les méthodes d'intervention sur les méthodes d'évaluation. Mais ces aspects sont étroitement liés. *Le chercheur qui souhaite appliquer une méthodologie rigoureuse de type expérimental va avoir tendance à limiter les objectifs et les modes d'interventions à des éléments simples et mesurables. Un éducateur pour la santé persuadé de l'importance de la participation des individus et de la population orientera probablement le choix de ses objectifs afin qu'ils soient cohérents avec ses modes d'intervention privilégiés.*

Les lignes qui suivent n'ont pas pour but de tenter une description, et encore moins une synthèse, des multiples formes d'éducation pour la santé et de promotion de la santé, mais simplement de montrer que cette diversité est une réalité incontournable pour tout évaluateur, réalité qu'il faut intégrer pour rendre compte de la diversité des pratiques d'évaluation en général et d'évaluation d'efficacité en particulier.

a) Le choix des objets de l'évaluation de l'efficacité.

Le manque de précision de la terminologie utilisée par les divers auteurs pour nommer les catégories d'objets à évaluer est significative de la difficulté de situer quels sont les buts de l'éducation pour la santé et de la promotion de la santé.

On trouve en effet sous les termes "*outputs*", "*outcomes*", "impact", des indicateurs assez différents, qui glissent d'une catégorie à l'autre. Les plus nombreux réservent le terme "*outcomes*" aux résultats épidémiologiques (Macdonald, 1997, par exemple). Ils situent alors les changements de comportements dans les "*outputs*". La notion d'"impact" est utilisée pour les effets globaux, à long terme (Pineault et Daveluy, 1986). Pour Green et al. (1980), par contre, la catégorie "impact" recouvre les changements directs liés à l'éducation pour la santé. Bury (1988) discute ces différences sémantiques, mais sans faire d'hypothèse sur leur origine.

Sans faire un inventaire précis, nous estimons que ces différences reflètent souvent de manière implicite une hiérarchie entre les objets à mesurer. Un examen plus approfondi dans ce sens pourrait se révéler éclairant. Nous préférons utiliser, de façon plus générale, le terme résultats ("*outcomes*") pour tous les effets attendus, qu'ils soient de nature cognitive, attitudinale, comportementale, épidémiologique ou autres (sans hiérarchie) et le terme effets pour tous les effets attendus et non attendus. Nous traduisons le terme "*outputs*" par réalisations, par référence aux ressources introduites dans le programmes ("*inputs*"), et nous n'utiliserons pas le terme impact.

Les résultats épidémiologiques.

Pour Nutbeam, Smith et Catford (1990), les études épidémiologiques établissent les besoins de santé et le lien causal entre les comportements des individus et les risques de santé. Les besoins de santé et les facteurs de risque sont des composantes essentielles à prendre en compte pour assurer avec succès une intervention d'éducation pour la santé.

La mortalité et la morbidité sont des indicateurs traditionnels de la santé. Cependant, les tentatives pour les utiliser en éducation pour la santé donnent des résultats peu concluants. En effet, cela supposerait l'existence d'un lien direct entre une intervention éducative et un changement de ces indicateurs (Nutbeam, Smith et Catford, 1990).

Tones (1992) met bien en évidence cette difficulté. Les indicateurs de résultats finaux en termes épidémiologiques sont impraticables. Quel lien établir entre une information dans une école sur la consommation modérée d'alcool et l'incidence de la cirrhose du foie dans la population lorsque ces jeunes seront devenus des adultes ? Il faut donc des indicateurs plus proches de l'intervention, comme la mesure de la rétention cognitive, par exemple.

Kok (1992) fait le même constat. Les résultats de l'éducation pour la santé en terme de santé sont souvent trop éloignés dans le temps pour être accessibles. Un compromis consiste à utiliser des indicateurs intermédiaires.

Les résultats comportementaux et cognitifs.

Les recherches en sciences sociales montrent les facteurs comportementaux sur lesquels il faut intervenir et les méthodes pour obtenir un changement. Les mesures comportementales et cognitives sont les plus courantes (Nutbeam, Smith et Catford, 1990).

Green (1989, p. 3) affirme très nettement l'efficacité de l'éducation pour la santé sur ce plan. *"Maintenant, la preuve de l'efficacité de l'éducation pour la santé est aussi forte que pour d'autres disciplines de la santé ou de l'éducation en ce qui concerne du moins le changement de comportement. Les méta-analyses réalisées dans divers champs d'action le montrent, surtout dans le secteur de l'éducation du patient. Il est vrai que les travaux y sont plus nombreux et sont plus adaptés à la méta-analyse."*

Les résultats attitudeaux, sociaux, environnementaux.

Pour évaluer l'efficacité d'une action, il est aussi nécessaire de construire des outils valides et fidèles qui mesurent des changements de valeurs, des changements sociaux ou des changements dans l'environnement (Nutbeam, Smith et Catford, 1990).

Il faudrait aussi évaluer les besoins de santé, les facteurs de risque et les facteurs psychosociaux, socio-économiques et environnementaux qui influencent ces facteurs de risque. Les difficultés sont alors dues à la longueur, à la complexité et à la multifactorialité de la chaîne de causalité.

Baric (1980) illustre ces difficultés par la notion de "carrière de santé" d'un individu. Pour lui, la "carrière de santé" est un processus complexe, permanent, qui ne peut être évalué de manière rigoureusement scientifique.

Deccache (1997, p. 11) fait la même distinction, mais n'aboutit pas à la même conclusion. Pour lui, *"le concept de santé est au moins double : d'une part, c'est une réalité épidémiologique, objectivée et mesurable, et d'autre part, c'est une réalité individuelle et sociale, subjective, mais tout aussi réelle et évaluable."*

Les résultats sur les processus de "gestion" de la santé.

Quels sont les indicateurs à identifier et construire pour rendre compte de l'effet d'une intervention sur la "carrière de santé" d'un individu ? Il s'agit ici d'influencer un processus. L'acquisition ou la modification d'un processus peut être un but et donc l'objet d'une évaluation d'efficacité. C'est peut-être même la voie la plus prometteuse pour le développement de l'éducation pour la santé et de la promotion de la santé. Pour permettre à l'individu de mieux gérer sa santé, il faudra créer les conditions qui favorisent l'apprentissage de compétences transversales, lui apprendre à apprendre.

La gestion de sa santé n'est pas l'application de préceptes de santé, mais la capacité de s'adapter à son milieu et de le modifier. Cette capacité d'adaptation passe par l'acquisition de connaissances, de comportements, mais surtout par l'aptitude à chercher et à analyser une information, à communiquer avec les autres, à s'affirmer, à faire des choix, à modifier volontairement des habitudes, etc. La promotion de la santé devra donc identifier, développer et évaluer ces processus de gestion de santé.

Les résultats sur la qualité de vie.

Que devient la chaîne de causalité si l'on envisage comme indicateurs de santé, non plus l'état de santé, mais la qualité de vie comme le suggère l'évolution de la recherche sur les résultats de santé commentée par Pransky et Himmelstein (1996) ?

L'introduction de la qualité de vie comme but de la promotion de la santé et de l'éducation pour la santé réoriente la question du choix des indicateurs et la problématique de la "chaîne" de causalité.

Par exemple, faut-il considérer qu'une intervention d'éducation pour la santé est inefficace si elle parvient à modifier positivement la perception de la santé, indicateur de qualité de vie, mais ne modifie pas l'état de santé biologique ?

Permettre à des individus, à des collectivités, d'avoir un haut niveau d'estime de soi, de se sentir responsables de leur propre vie, de se sentir capable de réaliser leurs objectifs (Green, Tones et Manderscheid, 1995, p. 107) sont-ils des objectifs intermédiaires ou des objectifs en eux-mêmes au regard de la qualité de vie ? Quel écart subsiste alors encore entre l'éducation et l'éducation pour la santé (Demarteau, 1985) ? Que devient l'évaluation de l'efficacité ?

C'est dans ce sens que Saan (1997, p. 35) conclut un article sur l'évaluation de la qualité : *"I hope quality in health promotion will seek to reflect in the long term the core value of health promotion itself : that health and quality are not an end or destination in themselves, but a journey. And the joy of life is only partly in arriving. It is mostly in the travelling, together. Quality in health promotion adds for me to the quality in life."*

Et si, de manière théorique peut-être, l'éducation pour la santé ne faisait qu'ajouter de la vie, de la qualité de vie, aux années sans ajouter d'années à la vie, n'atteindrait-elle pas néanmoins son but ?

Ainsi, le choix des indicateurs d'efficacité est bien tributaire des buts fixés à la promotion de la santé et à l'éducation pour la santé.

b) Les contraintes des méthodes d'intervention.

Au-delà de la complexité du choix des objets à évaluer, l'évaluation de l'efficacité est intimement liée aux modes d'intervention qui sont mis en œuvre. Ces moyens d'action conditionnent bien sûr les résultats, mais ils influencent aussi les méthodes d'évaluation.

A ce propos, Green (1977) décrit sept dilemmes posés par la nature de l'éducation pour la santé aux questions d'évaluation et de mesure.

La rigueur du schéma expérimental versus la pertinence et l'adaptabilité du programme.

La rigueur du schéma expérimental conduit à des interventions éducatives rigides et peu intéressantes. Une intervention en éducation pour la santé doit être souple et adaptée, mais comment décrire ces adaptations ? Les interventions significatives sont souvent décrites trop vaguement pour être répétées.

Mullen, Green et Persinger (1985) confirment l'importance de la pertinence et de l'adaptabilité par une méta-analyse des résultats de septante essais cliniques d'interventions éducatives auprès de patients chroniques. Le meilleur prédicteur de l'efficacité sur les connaissances et la compliance est un score décrivant la qualité éducative de l'intervention. Ce score comprend 6 critères :

- la cohérence (relation entre objectifs et méthodes),
- la pertinence (adéquation à l'expérience préalable des personnes),
- l'individualisation,
- le feedback,
- le renforcement,

- la facilitation de l'apprentissage.

On y reconnaîtra les critères classiques de la pédagogie "par objectifs" d'origine behavioriste (Bloom, 1979).

La validité interne ou efficacité "vraie" versus la validité externe ou faisabilité.

Comment et quand faut-il sacrifier l'une pour l'autre ? Pour Green (1977), la validité interne est plus importante dans une première étape puis, au moment d'envisager la généralisation, elle cède la place à la validité externe.

Nutbeam, Smith et Catford (1990) proposent la même procédure. Les auteurs suggèrent trois étapes clés à respecter lors du développement d'un programme. L'importance de l'évaluation de l'efficacité diminue progressivement et laisse la place à l'évaluation du processus.

- A partir de la recherche fondamentale, il y a d'abord les études expérimentales : est-ce que cela marche ?
- Puis viennent les études de démonstrations : est-ce que cela peut-être répété, généralisé ?
- Suivent enfin les études de dissémination : est-ce que c'est implanté ?

Dans la première étape, l'important est de vérifier les résultats. C'est l'étape qui intéresse le plus les chercheurs. Elle intéresse moins les décideurs et les praticiens (Nutbeam, Smith et Catford, 1990).

Dans la seconde, l'important est de comprendre pourquoi un programme fonctionne et sous quelles conditions. Cette étape est plus importante pour le praticien (Nutbeam, Smith et Catford, 1990).

Dans la troisième, l'évaluation a pour but de soutenir le programme. Elle se préoccupe de surveiller la qualité, de mesurer les coûts. C'est la plus utile pour le praticien. Elle est peu développée en promotion de la santé et en éducation pour la santé (Nutbeam, Smith et Catford, 1990).

Cette démarche, très rationnelle, est celle que nous avons présentée dans le cadre de l'évaluation en santé publique (Bréart et Bouyer, 1991). Elle a le défaut d'envisager l'intervention (le programme) comme un produit à mettre au point, à tester et à diffuser, à l'encontre de la remarque précédente sur la pertinence et l'adaptabilité. La mise au point d'un tel programme est souvent longue et coûteuse et prend peu en compte la variabilité des situations. Elle met l'éducateur pour la santé qui l'utilise en position de simple relais plus ou moins efficace en fonction de son respect du mode d'emploi. Cette approche "technologique" présente des avantages (notamment au niveau de la reproductibilité et des coûts d'utilisation), mais elle va à l'encontre d'autres stratégies éducatives et participatives.

Il faut donc envisager de pouvoir étudier et évaluer aussi chacune des étapes évoquées ci-dessus de manière indépendante : mettre au point directement en situation réelle des procédures susceptibles d'être généralisées ou étudier sur le terrain les pratiques qui se sont implantées "naturellement".

Les effets expérimentaux versus les effets placebo.

L'éducation pour la santé cherche à créer la confiance, à augmenter la croyance dans l'efficacité personnelle, à faire prendre conscience des bénéfices possibles, à faire participer et collaborer à l'intervention, à renforcer la désirabilité sociale de certains comportements. Ces stratégies vont à l'encontre de l'approche expérimentale qui essaye d'éliminer l'effet placebo, l'effet Hawthorne et la désirabilité sociale. L'éducation pour la santé est un "effet placebo organisé" (Green, 1977). Il y a une incompatibilité entre les deux démarches. La méthode d'intervention éducative ne peut faire l'objet d'un traitement expérimental.

L'efficience versus l'économie d'échelle.

Le dilemme consiste à créer des outils, des démarches qui restent individualisés, personnalisés, flexibles et adaptables, qui soient efficaces tout en diminuant les coûts par une standardisation et une diffusion large. Des solutions existent, telles que d'utiliser le groupe comme agent de changement plutôt que les interventions uniquement individualisées. On peut recourir aussi au support de bénévoles ou à des technologies de diffusion. Le choix de certaines stratégies uniquement sur base des coûts sans étude de l'efficacité ne mène à rien. Mais une étude coût-efficacité doit intégrer de nombreux paramètres. La difficulté est qu'en matière éducative, ce qui fonctionne dans un contexte ne fonctionne pas nécessairement dans l'autre (Green, 1977).

Le bénéfice versus le coût.

Il faut minimiser les coûts et maximiser les bénéfices. Mais souvent les interventions qui ont les bénéfices les plus élevés en termes de changement de comportement (individualisation, recherche des personnes à haut risque), sont aussi les plus coûteuses. Il faut adapter les interventions et leurs coûts au type de risque et au bénéfice escompté. Que faut-il privilégier dans l'évaluation ? (Green, 1977).

Les effets à long terme versus les effets à court terme.

En éducation pour la santé, l'apparition de la plupart des bénéfices demande du temps. Les effets immédiats sont souvent temporaires. D'autres effets, plus lents à s'établir, sont souvent plus stables. Il faut donc tenir compte de ces variations dans la mesure des effets : effet retard, effet déclinant, effet de rebond négatif, tendance historique, effet de contraste, etc. L'évaluation d'efficacité est confrontée à la durée (Green, 1977).

Le niveau d'investissement seuil versus les niveaux d'investissement à rendement décroissant des dépenses en éducation pour la santé.

Comme dans d'autres domaines, il y a en éducation pour la santé un seuil d'efficacité, un minimum requis pour obtenir un effet perceptible. Cependant il est difficile de définir quel est le minimum à mettre en œuvre, donc à investir pour atteindre l'objectif.

- Au-delà d'un point de saturation, le rendement des interventions devient décroissant.
- En cas d'investissement insuffisant, on peut constater une diminution de la sensibilité aux interventions malgré l'absence d'effet perceptible ; il faudra alors un investissement ultérieur plus important pour atteindre le seuil.
- Le temps entraîne une perte d'efficacité et la nécessité d'entretenir et de renouveler les interventions pour produire les effets (Green, 1977).

Malheureusement, les solutions proposées par Green (1977) restent limitées à la méthodologie. Il met l'accent sur le besoin de développer plus de données cumulatives et propose une standardisation des mesures, la répétition des expérimentations dans différentes situations, un meilleur compte rendu des interventions et une meilleure diffusion des expériences vers la pratique.

L'application de ces dilemmes à la promotion de la santé.

Ces dilemmes proviennent de la nature des méthodes de l'intervention éducative, mais ils peuvent être repris pour les démarches de promotion de la santé dans la mesure où la participation en constitue un axe majeur. Ils indiquent le besoin d'une réflexion approfondie sur l'évaluation d'efficacité.

Ainsi, Nutbeam, Smith et Catford (1990) soulignent l'impossibilité d'appliquer le schéma expérimental de l'essai contrôlé randomisé à la promotion de la santé.

Ce schéma est le modèle idéal pour l'évaluation de l'efficacité. Mais répondre à ses exigences limite l'évaluation à la mesure d'interventions très ciblées, sur un seul facteur, et dans des milieux "fermés" tels que l'école, l'hôpital, les lieux de travail, qui permettent un contrôle plus aisé des variables. En dehors de ces situations, le problème principal est celui de la randomisation.

Dans une communauté, il est virtuellement impossible de créer deux groupes équivalents tirés au hasard qui n'auraient pas de relation entre eux. Même en n'utilisant pas les médias ou les réseaux sociaux institués comme moyens d'intervention, les réseaux informels se chargent de faire circuler l'information.

La promotion de la santé et l'éducation pour la santé communautaire cherchent à atteindre la population, à modifier le cadre de vie et pas seulement le comportement d'un individu. L'échantillonnage aléatoire est inapplicable.

L'utilisation de schémas non expérimentaux, la sélection de communautés séparées comme unité de comparaison réduit le risque de contamination, mais diminue la possibilité de vérifier l'attribution des effets (Nutbeam, Smith et Catford, 1990).

4. Des remèdes.

Dans le cadre de leur analyse de la littérature sur l'efficacité, Veen et al. (1994) proposent une série de critères de qualité de la recherche évaluative en éducation pour la santé et promotion de la santé.

Pour répondre à la spécificité de ces disciplines, il faut assurer les conditions suivantes.

- Une formulation réaliste des objets de l'évaluation en fonction de la durée de l'intervention et des investissements ; une formulation suffisante des caractéristiques des bénéficiaires, de la nature du bénéfice, de son ampleur, de son moment.
- Une discussion des biais méthodologiques en fonction des contraintes de l'éducation pour la santé : adaptabilité de l'intervention, durée, diffusion des innovations, stratégies de communication ou de réseaux. L'utilisation d'instruments qualitatifs et quantitatifs qui permettent une validation croisée et la mise en œuvre de mesures répétées. La spécification des taux de réponses, des abandons, etc.
- L'évaluation des conditions de fonctionnement : stratégies, méthodes, contenu, perception des participants, raisons de ces réactions, conséquences imprévues. Cela permet de comprendre ce qui favorise ou entrave l'efficacité, cela améliore la compréhension des conditions de généralisation.

Ces propositions définissent la structure de la partie suivante : tout d'abord les réponses orientées vers le choix des objets, puis l'amélioration des méthodes existantes ainsi que le développement de nouvelles méthodes et enfin, l'intérêt croissant pour l'évaluation de processus.

D'autres solutions ont aussi été proposées pour dépasser les limites de l'évaluation de l'efficacité : elles prônent le développement simultané de la planification et de l'évaluation. Elles seront abordées dans le cadre du concept de la qualité.

a) Les objets.

Selon Nutbeam, Smith et Catford (1990), il faut d'abord des attentes raisonnées et raisonnables vis-à-vis des interventions et des évaluations. Attendre d'une intervention communautaire qu'elle produise des effets observables à long terme sur la morbidité et la mortalité est difficile, irréaliste et peu approprié. Il faut juger les programmes d'éducation pour la santé sur leur capacité à modifier les facteurs de risque, les comportements, les facteurs personnels, sociaux et environnementaux (Nutbeam, Smith et Catford, 1990).

b) Les méthodes.

On peut regrouper en quatre courants les propositions d'ordre méthodologique, prônées pour améliorer l'évaluation d'efficacité.

Se limiter à l'utilisation de données quantitatives.

En 1994, cinq directeurs de centres fédéraux de santé publique aux Etats-Unis publient un article réaffirmant la nécessité de mesurer l'efficacité par le recueil de données quantitatives sur les résultats de santé. Cette évaluation rigoureuse leur paraît essentielle pour convaincre de l'utilité de financer la prévention et la promotion de la santé (Thacker et al., 1994).

Développer les outils existants.

McQueen (1989) affirme que les méthodes pour mesurer les changements dans les attitudes et connaissances sont adéquates, mais pas celles pour mesurer les changements de comportements. Il faut développer des indicateurs plus fins à propos des styles de vie. Il faut surtout des méthodes de recueil de données plus dynamiques, continues, comme l'enquête téléphonique assistée par ordinateur CATI (*Computer Assisted Telephone Interviewing*).

Combiner les méthodes quantitatives et qualitatives.

De Vries et al. (1992) proposent d'utiliser l'approche qualitative pour générer des hypothèses qui seront ensuite vérifiées de manière quantitative. La triangulation, fondée sur le recoupement de différentes sources de données, apparaît aussi comme une solution pour valider les informations (de Haes, 1991, Macdonald, 1996).

De toute façon, il est irréaliste de n'adopter que le schéma expérimental pour l'évaluation des résultats. Il est en effet impossible de développer des interventions communautaires dans le cadre de ce schéma. Il faut concevoir et utiliser d'autres schémas, d'autres outils, d'autres démarches d'évaluation réalistes et appropriées (Nutbeam, Smith et Catford, 1990).

Développer de nouvelles approches.

Ces nouvelles approches ont été proposées pour rencontrer des situations complexes comme celles développées dans le cadre du mouvement des villes et villages en santé (*Healthy Cities*).

Dixon et Sindall (1994) situent l'enjeu de l'évaluation des programmes de promotion de la santé orientés vers la communauté (*community-oriented*). Il faut avant tout distinguer clairement les définitions ainsi que les buts internes et externes qui sous-tendent la notion de succès. Ils décrivent trois "logiques" de changement communautaire, qui appellent différentes méthodologies.

La première est d'origine interne et met en jeu toutes les valeurs sociales et tous les acteurs. Il faut alors développer une approche ethnographique, raconter l'histoire de la communauté.

La seconde est d'origine externe et se déroule en partenariat. C'est un programme communautaire. La démarche rationnelle tente de rencontrer les valeurs de la communauté. Les indicateurs et les critères de succès sont définis largement et négociés.

La troisième est aussi externe, mais basée sur une définition épidémiologique des besoins. Ce n'est plus un programme communautaire, mais une intervention communautaire. L'évaluation se préoccupe des résultats de l'intervention sur les facteurs de risque et le statut de santé, en utilisant une méthodologie coût-efficacité.

Hawe (1994) attire aussi l'attention sur les particularités des communautés et sur les différents sens que peuvent prendre les termes "implication communautaire", "développement communautaire", "participation communautaire". Une communauté n'est pas un ensemble d'individus et on n'évalue pas un programme d'intervention communautaire en mesurant individuellement les modifications de comportements et d'attitudes. Hawe propose de considérer le renforcement des capacités de gestion autonome de la communauté (*empowerment*) non comme un processus, mais comme un résultat à atteindre et à évaluer, par une démarche active et interactive de clarification des valeurs et des intentions.

Poland (1996) note qu'il y a encore peu d'évaluation systématique du mouvement des "villes en santé". Il identifie également la dimension éthique et politique que doit rencontrer l'évaluation. Il propose une série de principes pour guider l'évaluation des initiatives de "santé communautaire". Il met notamment en évidence l'approche participative, intégrée au programme ; en phase étroite avec celui-ci pour fournir des feedbacks rapides et compréhensibles ; centrée sur le processus, sensible aux valeurs multiples, utilisant diverses techniques. Il propose comme modèle conceptuel pour l'évaluation le modèle de Smith (1989). Ce modèle est basé sur l'identification des enjeux et des acteurs (*Stakeholders*).

On voit qu'au-delà de la diversification des méthodes, c'est à un véritable changement de paradigme que l'on assiste. Le paradigme endosystémique devient une méthode d'évaluation reconnue en promotion de la santé. Les modèles développés par l'évaluation en éducation, comme celui de l'évaluation par les enjeux, sont devenus attractifs.

c) Le processus.

Ainsi, les nouveaux modèles proposés s'orientent vers l'identification des critères de succès, vers l'analyse endosystémique. En outre, l'attention va se tourner autant vers le programme lui-même que vers ses effets. L'évaluation de processus devient une nouvelle priorité.

En 1989, un éducateur pour la santé de terrain, de retour d'une conférence internationale sur l'évaluation de l'efficacité, se plaignait du peu d'utilité de cette conférence pour sa pratique. Les programmes dont on lui prouvait l'efficacité n'étaient jamais décrits avec suffisamment d'informations pour pouvoir être reproduits (Lafontaine, 1990). Le seul moyen d'en savoir plus était d'abandonner les salles de conférence pour rejoindre le bar des congressistes.

Veen (1995) fait la même analyse de la pauvreté de la description des interventions dans la littérature internationale consacrée à l'efficacité et conclut à la nécessité de mieux décrire les interventions, de développer l'évaluation de processus, de se préoccuper de l'histoire du programme.

Cependant, comme dans toute forme d'évaluation, on trouve différentes manières de réaliser l'évaluation de processus en fonction de l'objectif poursuivi.

Pour certains, l'intérêt de l'évaluation de processus est lié à la compréhension de l'efficacité (Basch et al., 1985, Nutbeam, Smith et Catford, 1990, Veen, 1995).

L'évaluation de processus est surtout utile si un programme a montré qu'il atteint ses objectifs. Elle permet alors de savoir quelles activités ont produit quels résultats et dans quelles conditions, auprès de qui, avec quel effort. Elle permet d'identifier les liens de causalité (Nutbeam, Smith et Catford, 1990). Ces trois auteurs proposent d'étudier la dynamique de changement dans une communauté, de définir l'étendue et l'intensité de l'exposition à l'intervention, d'analyser l'acceptabilité, la perception des effets.

Basch et al. (1985) situent l'évaluation de processus dans le cadre théorique de l'étude de l'implantation. Ils restent dans une approche quantitative de vérification des effets pour éviter les erreurs d'interprétation de type III (attribuer un effet négatif à un programme qui n'a pas été correctement appliqué). Mais ces auteurs attribuent également une autre fonction à l'évaluation de l'implantation : la fonction formative, pour favoriser l'adaptation de programmes au contexte local.

La grille APEP permet d'identifier les différentes pratiques d'évaluation de processus introduites par ces auteurs.

Certaines pratiques mettent l'accent sur la composante jugement.

L'évaluation qui vise à contrôler l'utilisation des ressources humaines, matérielles, financières, temporelles selon le plan établi voit le processus comme l'enregistrement d'une série d'événements externes observables et mesurables.

Celle qui sert à favoriser l'adaptabilité et la pertinence voit le processus plutôt comme un mécanisme interne d'ajustement entre le programme et son environnement ; comme une négociation entre plusieurs valeurs.

D'autres pratiques permettent de "connaître" l'action avant de la juger.

C'est le cas de l'évaluation qui vise à interpréter les mesures quantitatives d'efficacité ou de celle qui vise à comprendre l'histoire du programme dans une perspective phénoménologique ou constructiviste.

D'autres pratiques enfin sont directement au service de la décision.

Ainsi, l'évaluation de processus peut être conçue comme le moyen de corriger, de renforcer ou comme l'occasion de réajuster, de modifier l'intervention.

En privilégiant une approche externe ou interne, formelle ou informelle de chacun de ces axes et en les combinant entre eux, la grille APEP permet d'identifier de multiples formes d'évaluation de processus.

III.C. COMMENT RESOUDRE LA QUESTION DE LA QUALITE ?

Dès le début de ce chapitre, nous avons lié l'évaluation d'efficacité à l'évaluation de la qualité en montrant le parallélisme des justifications apportées par différents auteurs à l'émergence de ces concepts et pratiques : effet de mode (Saan, 1997), climat économique, compétition pour les ressources (Macdonald, 1997, Ovreteit, 1996, Veen, 1995), raisons éthiques ou professionnelles.

En éducation pour la santé, l'évaluation de la qualité trouve son origine dans l'évolution de la réflexion sur l'efficacité. Face aux spécificités de l'éducation pour la santé et de la promotion de la santé dans le champ de la santé, certains auteurs vont reconnaître les limites du paradigme expérimental représenté par la recherche évaluative ; ils vont diversifier les objets de celle-ci, développer ses techniques, s'intéresser au processus et par là, rejoindre le concept de qualité, tout en conservant une vision quantitative.

Par ailleurs, le paradigme endosystémique s'approprie aussi ce concept, mais en y intégrant la compréhension de la dynamique des interventions.

Nous éclairerons tout d'abord le concept de qualité en apportant quelques éléments de réflexion sur ses origines dans les secteurs de l'entreprise et des services de santé. Après avoir précisé la différenciation à établir entre l'évaluation de la qualité et l'assurance de la qualité, nous poserons l'hypothèse d'une équivalence entre l'évaluation de la qualité et l'évaluation de programmes ou de services.

Sur cette base, nous tenterons ensuite une clarification terminologique à partir de deux articles récents (Macdonald, 1997, Deccache, 1997), en montrant différents niveaux d'interprétation du concept. Nous aborderons enfin la problématique de différentes modalités d'évaluation de qualité en prenant pour exemple l'évaluation de la qualité de la planification.

1. Le développement du concept de la qualité dans les entreprises.

Le terme "qualité" apparaît d'abord dans le champ de l'entreprise en relation avec la notion de contrôle de la qualité d'un produit, puis en relation avec le souci de la satisfaction du client, du consommateur.

Dans un document qui présente les normes correspondant à la gestion de la qualité de service dans le système ISO 9004 (*International Standard Organisation, 1991*), on peut lire que la qualité est une approche systématique dont le but est de garantir la compréhension et la satisfaction des besoins exprimés ou implicites du client.

Ce concept serait applicable à toutes les formes de service, pour des organismes de toutes les tailles. Il suppose l'existence d'un système de revue permanente de toutes les composantes du service et l'amélioration continue de ce dernier.

Initialement centré sur le processus industriel de production d'un produit pour en assurer le contrôle, le concept de qualité va, dans les entreprises, aboutir à la notion de qualité totale, c'est-à-dire la maîtrise de tous les aspects d'un cycle de production : le besoin ou le désir d'un client, le marché, la concurrence, la conception, la mise en œuvre des ressources, les procédures, la gestion du personnel, la culture d'entreprise, la réalisation du produit ou du service, la satisfaction du client, les ventes, et même, de plus en plus, les effets sur la société et l'environnement (ISO 9004, 1991, Cockerill Sambre, 1994).

Les techniques de gestion de la qualité ont connu une véritable explosion dans le secteur privé, qui s'explique par les bénéfices escomptés, mais aussi par la création d'un véritable marché de la gestion de qualité qui envahit tous les secteurs de l'activité (Kalisz, 1996).

Ce mouvement d'élargissement progressif a été illustré dans le chapitre 2, lors de la présentation des différents modèles d'évaluation en éducation.

Stake (cité par Stufflebeam et Shinkfield, 1985) propose un modèle d'évaluation qu'il qualifie de "*countenance model*" et qui s'inspire explicitement du concept de qualité totale, émergeant à l'époque dans les entreprises américaines.

Stufflebeam (1971) propose une vision élargie des objets de l'évaluation (CIPP). Sa réflexion, d'abord tournée vers les décisions d'implantation et de sélection ou de subvention, s'étend ensuite aux décisions de planification et aux décisions de structuration.

2. Le développement du concept de la qualité dans les services de santé.

Ovretveit (1996) met en évidence l'existence de différentes "approches" de la qualité qu'il va jusqu'à opposer. Il attire l'attention des responsables des programmes de promotion de la santé sur le besoin de choisir un modèle approprié à leurs objectifs et modes d'action.

Une première approche est celle du contrôle de qualité, appelé "Assurance de Qualité" aux Etats-Unis ou ESI (*External Standard Inspection*). Des normes sont définies, surveillées, et des actions sont entreprises quand ces normes ne sont pas rencontrées (Ovretveit, 1996).

La deuxième approche est celle du développement continu de la qualité (CQI - *Continuous Quality Improvement*) ou parfois gestion de la qualité totale (TQM - *Total Quality Management*). Elle considère tout travail comme un processus et incite les gestionnaires à analyser et améliorer ce processus avec des "outils de la qualité" (Ovretveit, 1996).

Ovretveit fait un rapide tour des débats qui opposent ces deux approches. L'approche ESI a rapidement envahi les services de santé parce qu'elle est plus simple à mettre en œuvre. Pourtant, elle présente plusieurs inconvénients.

Tout d'abord, la définition des normes souffre souvent d'un manque de rigueur, car elles s'appuient sur des hypothèses de causalité entre processus et résultats qui ne sont pas démontrées. Ce qui est déjà difficile pour les actions curatives l'est encore plus pour la promotion de la santé. Dans ce cas, atteindre la qualité dans le processus (formation, équipement, accessibilité, etc.) ne garantit pas des résultats. Il faut baser les normes sur des liens scientifiquement éprouvés (*evidence based*). Mais fonder le choix des normes sur des preuves scientifiques ne fait que renvoyer au débat sur l'efficacité. Comment prouver l'efficacité de la promotion de la santé et de l'éducation pour la santé ?

On peut ensuite déplorer le coût d'une démarche, qui fixe des normes relativement statiques. Une partie importante des ressources d'une équipe peut être engloutie par un système de qualité trop rigide. C'est pourquoi d'autres démarches ont vu le jour qui focalisent plus l'attention sur la description interne du processus que sur la fixation de normes externes. Cette simple description est souvent suffisante pour mettre en évidence des faiblesses et y remédier.

Nous ne nous attarderons pas sur cette apparente opposition entre les deux modèles. On y voit apparaître des caractéristiques déjà illustrées par ailleurs. D'une part, l'attention est portée sur le jugement de valeur et le contrôle externe. D'autre part, elle est mise sur la décision et la régulation interne du système. Nous avons aussi montré que la combinaison de ces diverses caractéristiques permet d'envisager des modalités d'évaluation encore nombreuses.

Au-delà de ces différences, Ovretveit met en avant deux principes importants dans toute référence au concept de la qualité : l'attention aux demandes du consommateur et la vision systémique. En conséquence, Ovretveit (1996) propose d'envisager trois aspects pour développer un programme de la qualité en promotion de la santé : le souhait des consommateurs, les priorités des professionnels et les nécessités de la gestion. C'est en quelque sorte la trilogie, bien connue en santé publique, des demandes, des besoins et des ressources. Ce qui manque peut-être à la promotion de la santé pour intégrer le concept de la qualité est une démarche systématique d'amélioration des pratiques.

3. L'équivalence entre l'évaluation de programmes ou de service et l'évaluation de qualité.

La proposition de Ovretveit marque la possibilité d'un rapprochement entre les démarches de qualité et de planification en santé publique. Ce n'est pas le fait du hasard. On peut établir un lien entre la démarche d'évaluation de la qualité et la planification en général en comparant leurs fondements méthodologiques. Les mêmes étapes, le même vocabulaire sont utilisés : cycle, définition des problèmes, etc.

Allant plus loin, nous défendons l'idée d'une équivalence entre la démarche d'évaluation et la démarche de qualité. L'une comme l'autre sont des processus qui permettent de comprendre, juger et réguler le cycle de l'action, depuis sa conception jusqu'à son aboutissement.

L'évaluation de programmes étend d'abord son attention à la variété des objets, puis s'intéresse au processus et ensuite à l'ensemble de la planification. L'évaluation de la qualité, d'abord orientée vers le processus, va s'intéresser aux ressources investies (intrants), aux résultats et enfin, au contexte en amont et en aval.

On retrouve donc cette unité fondamentale du processus d'action et du rôle de l'évaluation au sein de celui-ci. Si on accepte cette proposition d'équivalence entre l'évaluation et le processus de qualité, on pourra appliquer à ce dernier, la même grille de lecture que celle que nous proposons pour l'évaluation. On peut en effet envisager de multiples modalités à l'évaluation de la qualité, comme nous l'illustrons ci-dessous.

4. Une proposition : distinguer évaluation de la qualité et assurance de la qualité.

Identifier l'assurance de qualité comme une composante de l'évaluation de qualité paraît essentiel. L'évaluation de qualité peut avoir pour fonction explicite de porter un jugement. Elle sera alors nommée assurance de qualité.

Prouver la qualité du processus ou prouver l'efficacité (la qualité du résultat) ou encore prouver la qualité globale (de l'ensemble du programme, depuis la définition des objectifs à partir du contexte jusqu'aux effets attendus et non attendus) est une des fonctions de l'évaluation, qu'elle soit énoncée ou non, faite par des acteurs internes ou externes ou à l'intention de ceux-ci, formalisée ou non. Cette preuve devient une assurance de qualité, si elle est acceptée comme telle par les acteurs et les décideurs. Pour ce faire, elle devra, le plus souvent, être énoncée et formalisée.

De nombreux auteurs en promotion de la santé font, à juste titre, de l'évaluation le moyen de prouver et d'améliorer les pratiques (*to prove and to improve*), mais sans différencier les enjeux stratégiques et les méthodes de formalisation (Macdonald, 1997, Tones, 1992).

Deccache (1997) propose, fort pertinemment, de trouver une voie médiane entre l'intérêt d'utiliser des normes concrètes et externes et celui de tenir compte des caractéristiques spécifiques, souvent qualitatives, des programmes mis en place par des acteurs tout aussi concrets. Il faut effectivement tenter de combiner les avantages des uns et des autres en modulant les évaluations en fonction des objectifs qui leur sont assignés. Pour y arriver, il faut pouvoir utiliser plusieurs approches ou enrichir certaines pratiques en répondant aux besoins identifiés.

Pour le responsable d'un programme d'éducation pour la santé ou de promotion de la santé, l'enjeu sera d'identifier et de négocier les conditions de la preuve et d'utiliser les moyens pour l'administrer (assurer de la qualité). Il serait par exemple possible de réaliser une évaluation de la qualité totalement interne (avec un processus, un jugement et des décisions explicites et formels) et en plus, de la faire valider par une expertise externe menée par un spécialiste du processus et par un décideur. Une telle démarche tenterait de combiner la démarche de régulation avec celle de justification.

Un responsable de programme pourra aussi, s'il le souhaite, développer d'autres stratégies d'évaluation afin de construire un savoir à partir de l'action, en vue de le diffuser ou de l'utiliser pour prendre des décisions de régulation de l'action.

Ainsi, il faudrait avant tout mieux identifier les pratiques réelles et les conditions de leur apparition pour ne pas en rester à des propositions théoriques. C'est l'objet et l'enjeu de la recherche proposée dans la deuxième partie de ce travail.

5. Une clarification terminologique.

La proposition d'équivalence entre les concepts de qualité et d'évaluation permet une lecture critique de la littérature récente sur ce sujet.

L'examen de deux articles parus dans la revue "Promotion Education", consacrée au troisième congrès européen sur l'efficacité et la qualité de la promotion de la santé, montre le besoin d'un cadre de réflexion plus large pour situer les différentes formes d'évaluation et clarifier leur utilisation. Cette clarification est proposée à titre d'exemple et n'a pas la prétention d'être représentative de la littérature sur le sujet.

a) L'article de Deccache (1997) dénote une difficulté à percevoir la totalité de la démarche d'évaluation.

Deccache distingue assez nettement au départ l'évaluation d'efficacité de l'évaluation de qualité. L'évaluation d'efficacité vérifie si les objectifs visés sont effectivement atteints. L'évaluation de qualité s'intéresse au processus, à ce qui se passe "entre *l'input* et *l'output*". Cet auteur montre clairement les particularités de la promotion de la santé et leurs implications sur l'évaluation de l'efficacité et de la qualité. Il montre aussi comment la question du processus est devenue essentielle en promotion de la santé ainsi que le besoin d'élargir les objets et les méthodes. La qualité est ainsi présentée comme un moyen d'améliorer l'efficacité. La distinction entre évaluation d'efficacité et évaluation de qualité est, pour lui, d'abord une question de finalité. La première est sommative ; elle permet de juger pour ajuster dans le futur. La seconde est formative ; elle permet des ajustements immédiats en cours de programme.

Deccache ne distingue pas évaluation de qualité et assurance de qualité. Macdonald, par contre, introduit l'équivalence entre évaluation et évaluation de qualité.

b) Macdonald (1997) propose de joindre les indicateurs de qualité et d'efficacité.

Dans son article coexistent deux "conceptions" de la qualité.

La première, globale, inclut le processus, les résultats et même l'évaluation des besoins. Il présente un cycle de la qualité qui est en tout point superposable au cycle de la planification. La démarche de qualité précède, accompagne et clôture l'action. Cette présentation va dans le sens de notre proposition d'équivalence et de globalité.

On trouve par exemple "(...) a framework that allows us to measure the quality of interventions by assessing both the process management and the outcome effectiveness." (Macdonald, 1997, p. 6)

La seconde dissocie l'efficacité et la qualité tout en les rapprochant.

La qualité est plutôt associée à l'évaluation de processus. "*Process analysis through the use of TQM should help to identify the problem and facilitate a mechanism to rectify it.*" (Macdonald, 1997, p. 7)

En outre, dans ses conclusions, l'auteur propose une position intermédiaire. Il faut développer de façon intégrée des méthodologies de recherche évaluative et des outils de mesure de la qualité adaptés à la promotion de la santé. En effet, on retrouve dans ces deux démarches la nécessité de combiner évaluation du processus avec évaluation des résultats.

Il n'est dès lors pas étonnant que Macdonald (1997) fasse un rapprochement entre la nécessité de combiner les approches externe et interne de définition des critères de qualité et la tendance à combiner les méthodologies quantitatives et qualitatives dans la recherche évaluative.

La comparaison est tentante, mais les deux propositions ne portent pas sur le même plan. La première s'adresse au jugement de valeur et la seconde au processus qui y conduit.

Il nous semble que le glissement d'une conception à l'autre indique que le passage à une vision tout à fait globale de l'évaluation n'est pas encore terminé et que subsistent des désignations terminologiques qui confondent la fonction sociale, la position stratégique et les caractéristiques techniques.

Une approche avec des normes standardisées externes est centrée sur le jugement de valeur et la position stratégique externe, mais peut mettre en œuvre d'autres modalités assez proches en explicitant le processus et la décision. Une approche basée sur l'amélioration continue interne est plutôt orientée vers la décision avec une position stratégique interne avec les mêmes possibilités de variantes.

6. L'évaluation de la qualité de la planification.

L'étude de Van den Broucke et al. (1994) porte sur la qualité de la planification des programmes de promotion de la santé dans la Communauté flamande de Belgique. Deux raisons justifient de s'arrêter à cette initiative :

- c'est la seule étude que nous ayons identifiée qui traite des demandes de financement de programmes. L'étude utilise le même type de matériel que la nôtre ; ainsi, au-delà des différences méthodologiques dans l'analyse, les résultats obtenus de part et d'autre pourraient être comparés ;
- ce travail fournit l'occasion d'appliquer la grille APEP au concept de la qualité de la planification.

L'importance de la planification des programmes de promotion de la santé et d'éducation pour la santé apparaît assez tôt. Modolo et Mamon (1991) signalent sa présence dès 1962. Cependant, c'est surtout grâce à Green qu'elle a pris une place essentielle (Green et al., 1980, Green et Kreuter, 1991).

Parallèlement, l'évaluation joue un rôle essentiel dans une pratique raisonnée de l'éducation pour la santé et de la promotion de la santé, pour tester les premières décisions, améliorer l'intervention et vérifier l'efficacité.

Kok (1992) indique comment l'évaluation permet d'éviter les pièges spécifiques à chaque étape de planification :

- développer une intervention pour un problème qui n'existe pas ;
- viser un comportement qui n'a pas un lien clair avec le problème ;
- partir d'une idée fautive sur les déterminants du comportement ;
- choisir ou construire une intervention inadéquate ;
- négliger la phase d'implantation ;
- se satisfaire d'une intervention sans en évaluer les effets.

Kok (1992) va plus loin en faisant de la planification une garantie d'efficacité. Il s'appuie sur la méta-analyse de Mullen, Green et Persinger (1985) qui montre que, indépendamment des méthodes utilisées, la qualité de l'intervention éducative est liée à l'amélioration des connaissances et à la compliance de patients chroniques. Kok généralise ce résultat et fait de la planification en général le garant de toute activité de promotion de la santé.

La planification peut alors devenir à son tour un objet d'évaluation et un indicateur de la qualité des projets, comme le montre l'étude de Van den Broucke et al. (1994).

Ils partent du postulat que le succès d'un programme dépend de la qualité de sa planification en se référant, comme Kok, à l'étude de Mullen, Green et Persinger. Ils entreprennent alors de définir les indicateurs d'une bonne planification à partir de la littérature, construisent une grille d'analyse, vérifient la fidélité d'utilisation de cette grille par plusieurs personnes, puis interprètent les résultats obtenus pour juger de la qualité, proposer des remédiations et des critères de sélection. Lors d'une deuxième étape (Van den Broucke et Lenders, 1997), ils vérifient l'évolution de la qualité après deux ans.

Van den Broucke et al. discutent d'abord l'intérêt de l'approche par programmes. C'est une approche plus souple, plus focalisée sur la résolution de problèmes bien définis et qui permet de mieux adapter les ressources. Elle permet également de progresser en mesurant les succès et les échecs.

Ces auteurs nuancent cependant leur optimisme en reconnaissant que l'approche par programmes est surtout efficace par rapport au traitement de nouveaux problèmes. L'investissement nécessité par la planification la rend moins efficace pour les problèmes de routine. Ils reconnaissent aussi que beaucoup de projets n'offrent que l'apparence de programmes planifiés. C'est cette différence entre un modèle idéal de planification et la réalité qu'ils vont explorer en analysant 140 demandes de financement introduites auprès du gouvernement de la Communauté flamande.

Leur grille reprend 26 items répartis en 9 dimensions :

1. les données de départ vis-à-vis du problème de santé, des déterminants, des résultats attendus,
2. la définition du groupe cible,
3. la définition des objectifs,
4. la participation du groupe cible,
5. les méthodes proposées,
6. la planification,
7. l'évaluation de processus,
8. l'évaluation des effets,
9. la compétence de l'organisation.

Résultats de l'analyse des 124 programmes :

- En général, le problème de santé est bien documenté, mais pas les déterminants ni les résultats escomptés (85 % ne les mentionnent pas).
- La définition du public cible comprend des caractéristiques générales, l'identification du groupe et sa dimension globale, mais pas le nombre précis de personnes à atteindre ni les moyens de les toucher.
- La définition des objectifs correspond en fait à une formulation du niveau de prévention et des objectifs généraux. La spécification de l'accessibilité de ses objectifs est rare et la distinction entre objectifs principaux et intermédiaires encore plus (57 % ne le précisent pas).
- La participation du groupe visé est clairement indiquée pour 4 % des projets. Les méthodes sont relativement mieux définies (50 % de projets atteignent un niveau suffisant).
- La planification est également relativement mieux définie, mais 25 % des projets seulement ont un planning tout à fait clair.
- Si un quart des projets annoncent une évaluation du processus, ce processus n'est bien défini que dans 4 % des cas.
- Un quart des projets proposent une évaluation des effets qui comprend une mesure de départ et une mesure d'arrivée. Ils ne sont plus que 16 % à proposer une évaluation à moyen terme. Un seul projet propose des normes.

- La compétence de l'organisation est suffisamment précisée dans 59 % des cas.

Discussion des résultats par les auteurs :

Les auteurs expliquent en partie ces résultats par le contexte de l'appel d'offre qui n'a laissé que trois semaines de préparation aux organisations et dont les directives étaient fort vagues. D'autres déterminants sont la difficulté de disposer de données récentes et le contexte institutionnel qui a poussé bon nombre d'associations à profiter de l'opportunité offerte par l'appel d'offre pour financer des activités de services, alors que celles-ci ne se prêtent pas à une approche par projet.

Le manque de formation et d'expérience, en planification et en évaluation, sont aussi évoqués. Il paraît possible aux auteurs d'agir à ce niveau pour améliorer la qualité des projets et au-delà, la qualité des pratiques de promotion de la santé.

Pour estimer l'évolution dans la qualité de la planification, Van den Broucke et Lenders (1997) proposent une comparaison de ces résultats avec les résultats de l'analyse de 65 autres projets deux ans plus tard.

Ils notent une amélioration significative dans tous les domaines à l'exception de l'évaluation des effets. Les progrès les plus importants sont remarqués dans la précision des groupes cibles, des méthodes et de l'évaluation de processus.

Ils attribuent ces effets à différentes initiatives prises pendant les deux années qui séparent les deux analyses : formation, publication d'un guide sur la planification, précisions de l'appel d'offre quant aux procédures de sélection et aux critères de qualité.

Analyse de l'étude :

La démarche est intéressante, mais est normative et statique. Elle emprunte le concept de qualité basé sur des critères externes (ESI). Elle part du postulat du lien entre la qualité de la rédaction d'un projet, la qualité de la planification, la qualité du programme lui-même et des effets qu'il obtiendra, indépendamment des objectifs, des ressources investies et des activités entreprises. Ces hypothèses sont loin d'être démontrées. La méta-analyse de Mullen, Green et Persinger qui sert de "justification scientifique" ne permet pas d'établir ce lien de causalité (1985).

La norme de qualité est établie par des chercheurs travaillant dans l'organisme chargé de la coordination de la promotion de la santé dans la Communauté flamande et financé pour cela. L'appareillage scientifique mis en œuvre pour assurer de la fidélité de la codification intra-analyste et inter-analyste risque de masquer le jugement de valeur qu'implique l'établissement de ces normes, d'autant plus qu'elles peuvent ensuite servir à la sélection des projets, à un contrôle externe.

L'amélioration des résultats après deux ans peut tout simplement être le signe d'une plus grande habileté formelle à remplir des critères sans gain de qualité effectif dans la planification.

IV. CONCLUSIONS : LA FECONDITE DU MODELE

Dans le deuxième chapitre, nous avons présenté un grand nombre de modèles d'évaluation, nous avons montré que notre grille d'analyse permet de les décrire et de les classer en fonction de trois composantes principales et de trois axes de différenciation. Cette démarche fait apparaître la diversité des propositions, mais aussi leur unité de fond. Elle illustre la variété des enjeux en terme d'acquisition de connaissances sur l'action, de pluralisme des jugements et d'aide à la décision ainsi que les réponses apportées par les théoriciens de l'évaluation.

Le chapitre 3 a mis l'accent sur la description des pratiques d'évaluation par la recherche, en particulier sur l'utilisation des conclusions des évaluations et sur les activités des professionnels de l'évaluation. Il met en évidence l'influence variable de l'évaluation sur l'action en fonction du contexte et de la composante prioritaire utilisée : l'évaluation orientée vers la connaissance, le jugement ou la décision. Il montre aussi que les pratiques des évaluateurs ne sont pas unifiées, ni "modélisées", mais dépendent de leur contexte professionnel. Nous avons aussi souligné l'absence de l'évaluation "interne", conçue et menée par les gestionnaires des programmes.

Le chapitre 4 illustre comment la grille d'analyse APEP peut servir à étudier la permanence des composantes et des axes de différenciation de l'évaluation dans le domaine de la santé publique et de la promotion de la santé.

L'analyse de quelques cas montre la répétition des difficultés à identifier et conceptualiser les divergences rencontrées entre pratiques d'évaluation et modèles théoriques, à cause des enjeux que l'évaluation sous-tend. La conférence médicale de consensus que nous avons présentée en est un bel exemple. Mais ces difficultés s'expriment aussi dans des débats plus globaux sur la relation entre la connaissance et la décision en santé publique, sur la mesure de l'efficacité en santé au travail ou en éducation pour la santé, sur le rôle de l'accréditation des hôpitaux ou plus largement sur l'émergence du concept de "qualité" en promotion de la santé.

Nous montrons que le concept d'évaluation de la qualité pris dans son sens le plus large est le résultat d'une évolution qui permettrait de faire converger sur le terrain de la promotion de la santé deux logiques épistémologiques différentes, celle du paradigme expérimental et celle du paradigme endosystémique. L'élargissement des objets de l'évaluation aux besoins, aux intrants, aux processus, et plus seulement aux effets, met en avant la nécessité du pluralisme, de la polysémie, de l'apprentissage de l'action, tout en conservant la nécessité de fonder l'action sur des connaissances solides, de prouver sa qualité et ses effets, de prendre les meilleures décisions possibles en fonction des ressources et des circonstances. C'est le véritable enjeu d'une nouvelle évaluation, encore largement à inventer, pour dépasser les clivages toujours actuels entre les modèles épidémiologique et communautaire de la santé publique (Demarteau, 1994).

La grille APEP tente d'éclairer ces débats, ces enjeux, de reconnaître la diversité des situations, d'ouvrir des pistes de réflexion pour le développement de nouvelles modalités d'évaluation ou l'application de modalités existantes dans des contextes différents.

C'est le principal intérêt de cette grille d'analyse. Elle éclaire les enjeux des évaluations pour différents acteurs en fonction des situations rencontrées. Elle contribue par là à mieux rendre compte des possibilités et des limites des activités d'évaluation mises en œuvre pour développer la compréhension des actions, établir des normes et des critères de jugement ou aider aux décisions. A partir de cette clarification, elle devrait permettre de diversifier les pratiques d'évaluation et de les adapter aux besoins et demandes des acteurs, surtout aux acteurs directs des programmes, aux responsables de ceux-ci et à leurs bénéficiaires.

Il n'est peut-être pas nécessaire que tout intervenant maîtrise la réalisation d'enquêtes quantitatives ou le maniement d'outils statistiques et de schémas de recherche pour établir formellement une preuve de l'efficacité basée sur la construction d'un savoir scientifique. Par contre, ce praticien doit savoir que toute évaluation implique un jugement, il doit apprendre à négocier les "preuves" de la qualité de son travail et mettre en place les outils pour produire celles-ci. Il doit aussi voir l'évaluation comme un outil de réflexion et d'apprentissage de l'action, comme un moyen d'enrichir ses représentations et ses pratiques de l'action. L'évaluation peut l'aider à penser l'action.

La grille APEP met ainsi en évidence la nécessité d'une perception dynamique de l'évolution des pratiques d'évaluation sur le terrain pour permettre des interventions qui en développent la pertinence et l'utilité au service de la qualité générale des actions de promotion de la santé. Ces interventions peuvent consister en formations à l'analyse critique des pratiques actuelles et à la maîtrise d'une variété de modalités d'évaluation. Elle peuvent aussi prendre la forme d'un conseil aux évaluations internes et d'un suivi de celles-ci.

Affiner dans ce sens la description des pratiques d'évaluation des programmes d'éducation pour la santé dans la Communauté française de Belgique est le but de la recherche que nous présentons dans les chapitres suivants.

BIBLIOGRAPHIE

Alkin MC, Daillak R, White P. Using evaluations : does evaluation make a difference ? Beverly Hills: Sage, 1979.

Alkin MC, Ellett FS. An inquiry into the nature of evaluation theorizing. *Stud Educ Eval* 1979;5:151-6.

Alkin MC, Coyle K. Thoughts on evaluation utilization, misutilization and non-utilization. *Studies in educational evaluation* 1988;14:331-40.

Alkin MC (ed). *Debates on evaluation*. London: Sage, 1990:302.

Alkin MC, Ellett FS. Development of evaluation models. In: Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. London: Pergamon Press, 1990:15-21.

Alkin MC, Daillak R. Impact of evaluation studies. In: Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. London: Pergamon Press, 1990:106-9.

Barbier JM. *L'évaluation en formation*. Paris: PUF, 1990:307. 2è éd. (Pédagogies d'aujourd'hui).

Barbier JM. *Elaboration de projets d'action et planification*. Paris: PUF, 1991:287.

Baric L. Evaluation : obstacles and potentialities. *International Journal of Health Education* 1980;23(3):142-9.

Baric L. Evaluation : a tool for planning, management and auditing health promotion / education interventions. In: *Proceedings First European Conference on Effectiveness of Health Education*. Utrecht: Dutch Health Education Centre, 1989:23-27.

Basch CE, Sliepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding type III errors in health education program evaluation : a case study. *Health Education Quarterly* 1985;12(4):315-31.

Battista N, Tremblay J. Méthodes d'évaluation. In: Dufresne J, Dumont F, Martin Y. *Traité d'anthropologie médicale. L'institution de la santé et de la maladie*. Québec: Presses de l'Université, 1985:1245.

Baudier F. *Education pour la santé : guide pratique*. Besançon: Comité Départemental d'éducation pour la Santé, 1988.

Baumann M, Deschamps JP, Pechevis M, Tursz A. *Projets de recherche et mémoires en santé publique et communautaire. Guide pratique*. Paris: Centre International de l'Enfance, 1991.

Bloom BS. Caractéristiques individuelles et apprentissages scolaires. Bruxelles: Labor, 1989:270.

Borich GD. Decision-oriented evaluation. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:31-5.

Breart G, Bouyer J. Méthodes épidémiologiques en évaluation. Rev Epidém et Santé Publ 1991;39:5-14.

Broclain D. Comment s'élabore un consensus ? Oestrogènes et risque fracturaire, le cas de la conférence française sur la médicalisation de la ménopause. Santé Publique 1994;6(3):193-211.

Bryman A. The debate about quantitative and qualitative research : a question of method or epistemology ? Brit Journ of Soc 1984;35:75-90.

Bury JA. Education pour la santé : concepts, enjeux, planifications. Bruxelles: De Boeck Université, 1988:235.

Campbell DT, Stanley JC. Experimental and quasi-experimental design for research. Chicago: Rand Mc Nally, 1966.

Champagne F, Contandriopoulos AP, Pineault R. Cadre conceptuel à l'évaluation des programmes de santé. Rev Epid et Santé Publ 1985;33:173-81.

Chwalow J. Evaluation des programmes d'éducation pour la santé. Revue d'Epidémiologie et de Santé Publique 1991;39:83-87.

Clyne S. Adversary evaluation. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:77-9.

Cockerill Sambre. 1994, Qualité totale, objectif grand-angle. Liège: Cockerill Sambre, 1994.

Comte-Sponville A. Valeur et vérité : études cyniques. Paris : PUF, 1995.

Cook TD, Gruder CL. Meta-evaluation research. Eval Q 1978;2(1):5-51.

Cook TD, Reichardt CS. Qualitative and quantitative methods in evaluation research. Beverly Hills: Sage, 1979.

Cook TD, Campbell DT (eds). Quasi-experimentation : design analysis issues for field settings. Boston: Houghton Mifflin Company, 1979.

Cousins JB, Leithwood KA. Current empirical research on evaluation utilization. Review of Educational Research 1986;56:331-64.

- Cronbach LJ et al. Toward reform of program evaluation. San Francisco: Josey Bass, 1980.
- Crozier M, Friedberg E. L'acteur et le système. Paris: Seuil, 1977.
- Dab W. La décision en santé publique. Surveillance épidémiologique, urgences et crises. Rennes: ENSP, 1993:286.
- Danzon M. Apport de nouvelles techniques d'évaluation à la méthodologie de l'éducation pour la santé. Revue Internationale d'Education pour la santé 1980;20(3):27-34.
- Davidson PO. Evaluating lifestyle change programs. In: Davidson PO, Davidson SM. Behavioural medicine : changing health lifestyles. New York: Brunnet/Mazel, 1980:391-409.
- De Bock C, Demarteau M, Lonfils R. Bilan des programmes subventionnés. Education Santé 1993;82 (novembre):5-7.
- Deccache A. European seminar on evaluation in health education. Bruxelles: RESO Université Catholique de Louvain, 1988.
- Deccache A. Pour mieux choisir son évaluation... Définition et rôles des évaluations en éducation pour la santé. Liège: Association pour la Promotion de l'Education pour la santé, 1989:9.
- Deccache A. Evaluer la qualité et l'efficacité de la promotion de la santé : approches et méthodes de santé publique et de sciences humaines. Promotion and Education 1997;IV(2):10-4.
- De Haes WFM. Research and health education. Hygie 1991;X(2):13-5.
- De Ketele JM (ed). L'évaluation : approche descriptive ou prescriptive. Bruxelles: De Boeck-Westmael, 1986:288. (Pédagogies en développement. Problématiques et recherche).
- De Landsheere V, De Landsheere G. Définir les objectifs de l'éducation. Liège:Thone, 1975:293.
- De Landsheere G. Introduction à la recherche en éducation. Liège: Thone, 1976. 4è éd.
- De Landsheere G. Dictionnaire de l'évaluation et de la recherche pédagogique. Paris: PUF, 1979:338.
- De Landsheere G. Evaluation continue et examens ; précis de docimologie. Bruxelles: Labor, 1980.

De Landsheere G. La recherche expérimentale en éducation. Genève: Delachaux et Niestlé, 1982.

De Landsheere G. La recherche en éducation dans le monde. Paris: PUF, 1986.

Demarteau M, Pieron M. Analyse des communications verbales entre un professeur d'éducation physique et ses élèves. In: Pieron M (ed). Towards a science of teaching physical education : teaching analysis. Liège: A.I.E.S.E.P, yearbook, 1978;2:98-124.

Demarteau M, Simenya-Strauven C. Prédiction et évaluation formative : deux fonctions complémentaires ou contradictoires dans le cadre de l'opération PREDIC. LPE, Liège, octobre 1982.

Demarteau M. Les buts et l'action de la cellule scientifique de l'A.P.E.S. In: Actes de la journée d'études du 27 octobre 1984 de l'Association des Licenciés en Sciences Sanitaires de l'Université de Liège, Liège, 27 octobre 1984.

Demarteau M. Une approche pédagogique de l'éducation pour la santé, communication lors d'une journée de formation pour des enseignants des instituts supérieurs pédagogiques de l'enseignement officiel neutre subventionné, 26 novembre 1985.

Demarteau M, Reginster-Haneuse G. Comment les experts de l'Education pour la Santé peuvent-ils se mettre au service des organisations volontaires pour les aider à progresser vers la santé pour tous ? In: Actes de la XIIème Conférence Mondiale d'Education pour la Santé, Dublin, 1-6 septembre 1985. Dublin: Health Education Bureau, 1987:425-28.

Demarteau M, Reginster-Haneuse G. Bilan de cinq années de collaboration entre les universités et les acteurs de terrain dans la Communauté française de Belgique. In: Actes de la première Conférence Européenne d'Education pour la Santé, Madrid, 25-27 mars 1987. Madrid: Ministerio de sanidad y consumo, 1988:212-13.

Demarteau M. Quel débat éthique en éducation pour la santé ? Education Santé 1991;58:16-8.

Demarteau M. Guide pour les demandes de subvention. Education Santé 1992;68 (Suppl.):21-8.

Demarteau M. L'A.P.P.R.E.T. : une méthode pour construire des programmes d'éducation pour la santé. Liège: APES, 1992:27.

Demarteau M. Entre les pratiques de terrain et les pratiques de recherche, le rôle formatif des conseils en évaluation de l'A.P.E.S. auprès des personnels de santé et d'autres secteurs, communication à la 3ème journée scientifique R.E.S.O. - U.C.L., Bruxelles, 28 mars 1992.

Demarteau M. Le projet éthique de l'éducation pour la santé. In: L'Observatoire 1993;3:65-6.

Demarteau M. L'éducation pour la santé dans la démarche de santé publique : modèle épidémiologique ou modèle communautaire ? In: L'Observatoire 1994;3:25-31.

Demarteau M. Le conseil en évaluation : une fonction professionnelle à développer pour une évolution pratique de promotion de la santé, communication aux Troisièmes Rencontres Internationales francophones de Promotion de la Santé, Lille, 27-29 juin 1994.

Demarteau M, Vandoorne C. Le rôle de l'évaluation dans le programme d'amélioration de l'environnement nutritionnel à l'école "je mange bien à l'école". Partage 1994;3(3):6-8.

Desrosières A. La politique des grands nombres, histoire de la raison statistique. Paris: La Découverte, 1993:441.

De Vries H, Weijts W, Dijkstra M et al. The utilization of qualitative and quantitative data for health education program planning, implementation and evaluation : a spiral approach. Health Education Quarterly 1992;19(1):101-15.

Dignan MB, Carr PA. Program planning for health education and health promotion. Philadelphia: Lea & Febiger, 1987.

Dignan MB. Measurement and evaluation of health education. Springfield: Charles C. Thomas, 1989.

Direction générale de la santé. Promotion de la santé : méthodologie. Symposium, Paris, 1985. Paris: La Documentation Française, 1986.

Dixon J, Sindall C. Applying logics of change to the evaluation of community development in health promotion. Health Promotion International 1994;9(4):297-309.

Dorr-Bremme DW. Naturalistic evaluation. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:66-8.

Education Santé. Les subventions aux programmes d'action en éducation pour la santé. Education Santé 1989;38 (mai):2-7.

Eisner EW. Educational connoisseurship and criticism : their form and functions in educational evaluation. Journal of Aesthetics Education 1976;3-4:135-50. Reprinted in: Madaus GF, Scriven M, Stufflebeam DL. Evaluation models, viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhof, 1983:335-47.

Eisner EW. The perceptive eye : toward the reformation of educational evaluation. California: Stanford Evaluation Consortium, 1975.

- Floden RE. Flexner, accreditation, and evaluation. In: Madaus GF, Scriven M, Stufflebeam DL. Evaluation models, viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhof, 1983:261-77.
- Gatherer A, Parfit J, Porter E, Vessey M. Is health education effective ? : Abstracts, bibliography and overview of evaluated studies. London: The Health Education Council, 1979:92.
- Glanz K, Lewis FM, Rimer BK (eds). Health behavior and health education : theory, research and practice. San Francisco: Jossey Bass, 1990.
- Green LW. Evaluation and measurement : some dilemmas for health education. *AJPH* 1977;67(2):155-61.
- Green LW et al. Health education planning : a diagnostic approach. Palo Alto: Mayfield, 1980.
- Green LW, Lewis FM. Measurement and evaluation in health education and health promotion. Palo Alto: Mayfield, 1986.
- Green LW. The effectiveness of health education. In: Proceedings First European Conference on Effectiveness of Health Education. Utrecht: Dutch Health Education Centre, 1989:3-4.
- Green LW, Kreuter MW. Health promotion planning : an educational and environmental approach. Mountain view: Mayfield, 1991:506.
- Green J, Tones K, Manderscheid JC. Efficacité et utilité de l'éducation à la santé à l'école. *Revue française de Pédagogie* 1995;114:103-20.
- Goldenhar LM, Schulte PA. Intervention research in occupational health and safety. *JOM* 1994;36(7):763-75.
- Gremy F, Goldberg M. Les sciences de l'information dans le domaine de la santé. Paris: Faculté de médecine, 1981:325.
- Guba EG, Lincoln YS. Effective evaluation. San Francisco: Josey Bass, 1981.
- Guba EG, Lincoln YS. Fourth generation evaluation. London: Sage, 1989.
- Hawe P. Capturing the meaning of "community" in community intervention evaluation : some contributions from community psychology. *Health Promotion International* 1994;9(3):199-210.
- House ER. Assumptions underlying evaluation models. *Educational Researcher* 1978;7(3):4-12.

House ER. How we think about evaluation. In: House ER (ed). New directions in educational evaluation. London: Falmer Press,1986:31-47.

House ER (ed). New directions in educational evaluation. London: Falmer Press, 1986:261.

ISO. ISO 9004-2. Genève: Organisation Internationale de Normalisation, 1991.

Iverson BK, Levy SR. Using meta-analysis in health education research. Journal of School Health 1982:234-39.

Kalisz S. A quelle sauce, les travailleurs seront-ils "gérés" ? Le Ligueur 1996;46:17.

Keeves JP. Knowledge Diffusion in Education. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:715-22.

Kok G. Research in order to support health education in practice. In: Proceedings First European Conference on Effectiveness of Health Education. Utrecht: Dutch Health Education Centre, 1989:9-14.

Kok G. Quality of Planning as a Decisive Determinant of Health Education Effectiveness. Hygie 1992;XI(4):5-9.

Lafontaine C. L'efficacité de l'éducation pour la santé : une première conférence européenne. Education Santé 1990;45:2.

Latour B. Nous n'avons jamais été modernes. Essai d'anthropologie symétrique. Paris: La Découverte, 1991:211.

Leclercq D. Conception d'intervention et construction de produits de formation. Liège: Service de Technologie de l'Education de l'Université de Liège, 1995.

Ledwith F. Can Evaluation be cost effective or even effective ? Health Education Research 1986; 1(4):295-8.

Leviton L, Hughes E. Research on the utilization of evaluations : a review and synthesis. Evaluation Review 1981;5(4):525-48.

Lewy A. Formative and Summative Evaluation. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:26-8.

Liedekerken PC, Jonkers R, de Haes WFM, Kok GJ, Saan JAM. Effectiveness of health education. Assen: Van Gorcum, 1990.

Macdonald G, Adams L. A code of conduct for the Society of Health Education Officers. Journal of the Institute of Health Education 1986;24(3):111-5.

Macdonald G. Where next for evaluation ? *Health Promotion International* 1996;11(3):171-3.

Macdonald G. Quality indicators and health promotion effectiveness. *Promotion and Education* 1997;IV(2):5-8.

Madaus GF, Scriven M, Stufflebeam DL. *Evaluation Models*. Boston: Kluwer Nijhoff, 1983.

Madaus GF, Stufflebeam DL, Scriven MS. Program evaluation : a historical overview. In: Madaus GF, Scriven MS, Stufflebeam DL. *Evaluation models : viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff, 1983:423.

Marsch DD, Newman WB, Bover WF. Comparing ideal and real. A study of evaluation practice in California using the Joint Committee's Evaluation Standards. Los Angeles: AERA, 1981.

McQueen D. Indicators for determining the effectiveness of health education. In: *Proceedings First European Conference on Effectiveness of Health Education*. Utrecht: Dutch Health Education Centre, 1989:5-8.

Moatti JP. Efficience, vous avez dit efficience ? *Sciences sociales et Santé* 1992;X(1):125-33.

Modolo MA, Mamon J. The contribution of the IUHE to the evolution of key areas in health education : forty years in perspective. *Hygie* 1991;X(2):7-11.

Monnier E. *Evaluations de l'action des pouvoirs publics : du projet au bilan*. Paris: Economica, 1987:169.

Monoyer M, Miermans MC, Demartean M. L'approche judiciaire en évaluation. *Education Santé* 1994;88:7-9.

Mullen PD, Green LW, Persinger GS. Clinical trials of patient education for chronic conditions : a comparative meta-analysis of intervention types. *Preventive Medicine* 1985;14:753-81.

Nevo D. The conceptualization of educational evaluation : an analytical review of the literature. In: House ER (ed). *New directions in educational evaluation*. London: Falmer Press, 1986:15-29.

Nguyen-Nam T. L'éducation pour la santé en Belgique francophone (1985-1988). *Regard sociologique sur un champ socio-sanitaire émergent*. *Education Santé* 1990;46:5-10.

Nioche JP. De l'évaluation à l'analyse des politiques publiques. *Revue française de science politique* 1982;32:32-61.

- Nutbeam D. Communication between the uses and producers of research findings. Report on a WHO Seminar, Linköping, Sweden, 15-17 June 1987. Copenhagen: OMS Europe, 1987:15.
- Nutbeam D, Smith C, Catford J. Evaluation in health education : a review of progress, possibilities and problems. *Journal of Epidemiology and Community Health* 1990;44(2):83-9.
- Ouslikh M, Moulin M, Herscovici A. Les pratiques en éducation pour la santé. Analyse d'une professionnalisation en cours. In: Moulin M, Poucet T (ed). *L'éducation pour la santé, quelle(s) éthique(s) ?* Bruxelles: Centre de Diffusion de la Culture Sanitaire, 1993:143-66.
- Ovretveit J. Quality in health promotion. *Health Promotion International* 1996;11(1):55-62.
- Owens TR. Educational evaluation by adversary proceedings. In: House ER. *School evaluation : the politics and process*. Berkeley: Mc Cutchan, 1973.
- Pame P. Evaluation et promotion de la santé : état de la question à travers la littérature internationale. *Promotion and Education* 1995;II(1):37-40.
- Parlett MR. The new evaluation. *Trends in education* 1974;34:13-8.
- Parlett MR, Hamilton D. Evaluation as illumination : a new approach to the study of innovatory programs. Occasional Paper n°9. University of Edinburg, Center for Research in Educational Sciences 1972. Reprinted in: Glass GV (ed). *Evaluation studies review annual*. New York: Sage, 1976;1.
- Parlett MR. Illuminative evaluation. In: Walberg HJ, Haertel GD. *The International Encyclopedia of Educational Evaluation*. London: Pergamon Press, 1990:68-73.
- Patton M. *Utilization-focused evaluation*. Londres: Sage, 1978.
- Patton M. *Creative evaluation*. Beverly Hills: Sage, 1981:296.
- Patton M. The evaluator's responsibility for utilization. In: Alkin MC (ed). *Debates on evaluation*. London: Sage, 1990:185-202.
- Piette D. La professionnalisation des éducateurs pour la santé en Communauté française de Belgique. *Education Santé* 1990;52:3-12.
- Pineault R, Daveluy C. *La planification de la santé : concepts, méthodes, stratégies*. Québec: Agence d'ARC, 1986:480.

- Poland BD. Knowledge development and evaluation in, of and for healthy community initiatives. Part I : guiding principles. *Health Promotion International* 1996;11(3):237-47.
- Pransky G, Himmelstein J. Outcomes research : implications for occupational health. *American Journal of Industrial Medicine* 1996;29:573-83.
- Rimer B, Keintz M, Glassman B, Kinman J. Health education for older persons : lesson from research and program evaluations. *Advances in Health Education* 1986;1:369-96.
- Rootman I. Continuous quality improvement in health promotion : some preliminary thoughts from Canada. *Promotion and Education* 1997;IV(2):23-5.
- Rossi PH, Freeman HE. *Evaluation : a systematic approach*. Londres: Sage, 1982.
- Saan H. Quality revisited. *Promotion and Education* 1997;IV(2):34-5.
- Sanson-Fisher R et al. Developing methodologies for evaluating community-wide health promotion. *Health Promotion International* 1996;11(3):227-36.
- Scheerens J. Beyond decision-oriented evaluation. In: Walberg HJ, Haertel GD. *The International Encyclopedia of Educational Evaluation*. London: Pergamon Press, 1990:35-40.
- Schmidt FI. What do data really mean ? Research findings, meta-analysis and cumulative knowledge in psychology. *American Psychologist* 1992;47(10):1173-1181.
- Scriven M. The methodology of evaluation. In: Stake RE (ed), *American educational research association monograph on evaluation. curriculum evaluation*. Chicago: Rand Mc Nally, 1967;1.
- Scriven M. Pros and cons about goal-free evaluation. In: Hamilton D, Jenkins D, King C, MacDonald B, Parlett M (eds). *Beyond the number game : a reader in educational evaluation*. London: Macmillan, 1977:130-31.
- Scrivens E. Assessing the value of accreditation systems. *European Journal of Public Health* 1997;7:4-8.
- Sensi D. *Evaluer des projets d'innovation en éducation*. Bruxelles: Labor, 1990:133. (Education 2000).
- Sensi D. *L'évaluation des programmes d'éducation à la commission européenne : une approche comparative [Thèse de doctorat en sciences de l'éducation]*. Liège: Université de Liège, 1994:258.
- Shadisch WR, Epstein R. Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. *Evaluation Review* 1987;11:555-90.

Simeya C, Demarteau M. Les attitudes des enseignants devant l'évaluation. Comment les mesurer ? *Scientia Paedagogica Experimentalis* 1984;XXI (2):237-47.

Smith N. The feasibility and desirability of experimental method in evaluation. *Evaluation and Program Planning* 1980;3(4):251-5.

Smith MF. *Evaluability assessment : a practical approach*. Boston: Kluwer Academic 1989.

Stecher B. Goal-free evaluation. In: Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. London: Pergamon Press, 1990:41-2.

Stengers I. *Sciences et pouvoirs. La démocratie face à la technoscience*. Paris: La Découverte, 1997:120.

Straw RB, Cook TD. Meta-evaluation. In: Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. London: Pergamon Press, 1990:58-60.

Stufflebeam DL. *Educational evaluation and decision making*. Itasca: Peacock, 1971.

Stufflebeam DL. *L'évaluation en éducation et la prise de décision*. Ottawa: Editions NHP, 1980.

Stufflebeam DL, Webster WJ. An analysis of alternative approaches to education. *Educational Evaluation and Policy Analysis* 1980;2(3):5-20.

Stufflebeam DL, Shinkfield AJ. *Systematic evaluation: a self-instructional guide to theory and practice*. Boston: Kluwer-Nijhoff Publishing, 1985:347.

Suchman EA. *Evaluative research, principles and practice in public service and social action program*. New York: Russel Sage Fondation, 1967.

Thacker SB et al. Assessing prevention effectiveness using data to drive program decisions. *Public Health Reports* 1994;109(2):187-94.

Tones K. Health education and the ideology of health promotion : a review of alternative approaches. *Health Education Research* 1986;1(1):3-12.

Tones K, Tilford S, Robinson Y. *Health education : effectiveness and efficiency*. London: Chapman and Hall, 1990:286.

Tones K. Measuring success in health promotion : selecting indicators of performance. *Hygie* 1992;XI(4):10-14.

Touraine A. *Sociologie de l'action*. Paris: Seuil, 1966.

- Tyler R, Cagne R, Scriven M (eds). *Perspective of curriculum evaluation*. Chicago: Rand Mc Nally. (Monograph series on curriculum evaluation; 1).
- Van den Broucke S, De Potter B, Lenders F, Vinck J. An assessment of the methodological quality of health promotion projects in Flanders. *Arch Public Health* 1994;52:261-77.
- Van den Broucke S, Lenders F. Monitoring the planning quality of health promotion projects in Flanders. *Promotion and Education* 1997;IV(2):26-8.
- Vandoorne C, Demarteau M, Kefer C. L'évaluation, clé de l'évolution du programme Réseau Actions Sécurité. *Education Santé* 1992;70:14-20.
- Veen CA, Vereijken CJM, Driel WG van, Beliën MARE. Un outil d'analyse des études d'efficacité de la promotion de la santé et de l'éducation pour la santé : mise au point, utilisation et recommandations. Utrecht: Landelijk Centrum GVO, 1994:69.
- Veen CA. Evaluation of the IUHPE-project on the effectiveness of health promotion and health education. Proceedings of the expert meeting, Bonn, 22-24 March 1995. Utrecht: IUHPE regional Office for Europe, 1995:77.
- Vial MP. Evaluer n'est pas mesurer. *Cahiers pédagogiques* 1987:256.
- Walberg HJ, Haertel GD. *The international encyclopedia of educational evaluation*. London: Pergamon Press, 1990:796.
- Weber M. *Le savant et le politique*. Paris: Plon, 1959.
- Weiss CH. The stakeholder approach to evaluation : origins and promise. In: House ER (ed). *New directions in educational evaluation*. London: Falmer Press, 1986:145-57.
- Weiss CH. Toward the future of stakeholder approaches in evaluation : origins and promise. In: House ER (ed). *New directions in educational evaluation*. London: Falmer Press, 1986:186-200.
- Weiss CH. Evaluation for decisions : is anybody there ? Does anybody care ? In: Alkin MC (ed). *Debates on evaluation*. London: Sage, 1990:171-84.
- Weiss CH. If program decisions hinged only on information : a response to Patton. In: Alkin MC (ed.). *Debates on evaluation*. London: Sage, 1990:208-22.
- Williams JE. A numerically developed taxonomy of evaluation theory and practice. *Evaluation Review* 1989;13(1):18-31.
- Windsor RA, Baranowski T, Clark N, Cutter G. *Evaluation of Health Promotion and Education Programs*. Palo Alto: Mayfield, 1984.

Wolf RL. Judicial evaluation? In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:79-81.

Worthen BR. Program evaluation. In: Walberg HJ, Haertel GD. The international encyclopedia of educational evaluation. London: Pergamon Press, 1990:42-7.

TABLE DES MATIERES

Remerciements	2
Sommaire	3

INTRODUCTION

ENJEUX DE L'EVALUATION	5
L'évaluation sert à connaître l'action	5
L'évaluation sert à juger l'action	6
L'évaluation sert à guider l'action	6
La triple fonction de l'évaluation	6
ORIGINE DE LA RECHERCHE	8
Les premiers pas d'un parcours scientifique	8
L'expérience en éducation pour la santé en Communauté française	8
La pratique du conseil en évaluation	9
Un apport original	9
OBJECTIF ET METHODE DE LA RECHERCHE	11
Questions de départ	11
Objectif de la recherche	12
Objets de l'étude	12
Démarche méthodologique	12

CHAPITRE 1 LA GRILLE APEP : ANALYSE DES PRATIQUES D'EVALUATION DE PROGRAMMES

I. UNE DEFINITION DE L'EVALUATION	16
II. LES AXES DE DIFFERENCIATION DES PRATIQUES D'EVALUATION	18
II.A. La dimension sociale ou l'axe implicite – explicite	18
II.B. La dimension stratégique ou l'axe interne – externe	20
1. Le processus d'évaluation	24
2. Le jugement de valeur	25
3. L'utilisation de l'évaluation	26
II.C. La dimension technique ou l'axe informel – formel	29

1. La formalisation du processus d'évaluation	30
a) Délimiter les informations à recueillir	30
b) Obtenir les informations souhaitées par une démarche formelle et des procédures d'analyse	31
c) Fournir les informations	31
d) L'enjeu social de la formalisation du processus d'évaluation	31
2. Le jugement de valeur	32
3. La formalisation de l'utilisation de l'évaluation	33
III. LA GRILLE D'ANALYSE APEP	34
III.A. Les huit types d'évaluation	34
III.B. Les 124 modalités d'évaluation de la grille d'analyse	37
IV. LA DESCRIPTION DES PROGRAMMES	39
IV.A. La structure d'un programme	40
1. Le modèle de Barbier : le procès de conduite des actions	40
2. Le modèle CIPP de Stufflebeam	41
a) L'évaluation de contexte contribue à déterminer les objectifs	42
b) L'évaluation d'intrant aide à déterminer les structures des projets	42
c) L'évaluation de processus aide à contrôler les opérations	43
d) L'évaluation de produit aide à juger et à réviser les projets	43
3. Les catégories d'analyse de la structure d'un programme	43
IV.B. Le contenu d'un programme	44
1. Les objectifs et objets de l'évaluation de contexte et de produit	45
2. Le plan et les objets de l'évaluation d'intrant et de processus	46
V. LA THESE	47

CHAPITRE 2 LES MODELES ET PARADIGMES D'EVALUATION

I. L'EVALUATION "DECISION"	50
II. L'EVALUATION D'EXPERTISE	53
II.A. La classification de House	54
II.B. Le paradigme d'expertise	56
II.C. Le paradigme de prétoire	57
II.D. Les pratiques de méta-évaluation	58
III. L'EVALUATION FORMELLE	62
III.A. L'évaluation comme production de connaissance	62
III.B. L'évolution des paradigmes de l'évaluation scientifique	63
1. Le paradigme statistique	63

2. Le paradigme expérimentaliste	64
a) L'émergence d'une organisation scientifique du domaine social	64
b) L'impulsion de Tyler dans le domaine de l'évaluation	65
c) L'explosion de l'évaluation dans les années soixante	66
d) L'approfondissement des méthodologies de la recherche évaluative	68
3. Le paradigme endosystémique	69
4. La primauté du paradigme expérimentaliste	70
III.C. Les modalités de l'évaluation formelle	72
IV. L'EVALUATION DE CONTROLE	74
IV.A. Le paradigme juridique	74
IV.B. Le paradigme technico-économique	75
IV.C. Les liens avec l'évaluation d'expertise	75
IV.D. Les modalités de l'évaluation de contrôle	76
V. L'EVALUATION ORIENTEE VERS LES CONCLUSIONS	78
V.A. Scriven	78
1. L'évaluation indépendante des buts	79
2. L'évaluation formative et sommative	80
V.B. Stake	81
1. Principes du modèle de la "contenance"	81
a) La description et le jugement	81
b) Les antécédents, transactions et résultats	82
c) La comparaison	82
d) Les normes	82
2. Principes du modèle "répondant"	82
a) L'approche "répondante" est subjectiviste	83
b) L'approche "répondante" est pluraliste	83
c) L'évaluation "répondante" n'est pas par nature une recherche naturaliste ou qualitative	83
V.C. Cronbach	84
1. L'évaluation et l'évaluateur font partie de l'action politique	84
2. L'évaluation doit être compréhensive plus qu'explicative	84
3. La communication des résultats de l'évaluation est une tâche importante et délicate	85
4. L'évaluation doit guider et éclairer plus que juger, car il n'y a pas de vérité universelle	85
5. L'évaluation doit éclairer l'action sociale, mais par l'ensemble de ses interventions et non vis-à-vis d'un programme en particulier	85
V.D. Les idéologies de l'évaluation	86
1. L'idéologie séparatiste	86
2. L'idéologie positiviste	86
3. L'idéologie managériale	86
4. L'idéologie relativiste	87
5. La position idéologique de Scriven	87
V.E. Les modalités de l'évaluation orientée vers les conclusions	88
1. La première forme est interne au niveau du processus et du jugement	88
2. La seconde forme est interne au niveau du processus et externe	

au niveau du jugement	89
3. La troisième forme est externe au niveau du processus et interne au niveau du jugement	89
4. La quatrième et dernière forme prévue est entièrement externe	89
VI. L'EVALUATION "PROCESSUS - DECISION"	90
VI.A. Les distinctions entre l'évaluation formative et sommative	91
VI.B. Les différents modèles théoriques et l'évaluation orientée vers la décision	92
VI.C. L'approche "décisionnelle" de Stufflebeam	94
VI.D. L'approche "centrée sur l'utilisation"	95
VI.E. L'approche "par les enjeux"	97
1. L'analyse de Weiss	97
a) Les groupes des personnes qui tiennent les enjeux peuvent être identifiés avant le début de l'évaluation	98
b) Les personnes concernées souhaitent une évaluation à laquelle ils seraient associés	98
c) Ils souhaitent des informations particulières pour les aider à planifier et à faire des choix	98
d) Les évaluateurs peuvent répondre aux demandes formulées	98
e) Les participants à l'évaluation développent un sentiment de propriété envers les conclusions	98
f) Les décideurs parmi les participants (politiques, administrateurs, financeurs, directeurs de programme) utilisent les résultats comme base de décision	99
g) Même si les résultats ne mènent pas à des décisions spécifiques, ils enrichissent les discussions sur le futur du programme et éclairent les acteurs	99
2. Le débat de Weiss et Patton	100
VI.F. Les modalités de l'évaluation "processus - décision"	101
1. La première forme est interne tant pour le processus que pour la décision	102
2. La deuxième forme contient un processus interne et une décision externe	102
3. La troisième forme prévoit un processus externe et une décision interne	102
4. La quatrième forme est entièrement externe	102
VII. L'EVALUATION TOTALEMENT EXPLICITE	103
VII.A. Evolution de la définition de l'évaluation de Stufflebeam	103
VII.B. Monnier	103
VII.C. Vial	103
VII.D. Les modalités de l'évaluation explicite	105

CHAPITRE 3

LES RECHERCHES SUR L'UTILISATION ET LA PRATIQUE DE L'EVALUATION

I. L'UTILISATION DE L'EVALUATION

I.A. Patton (1978) : l'utilisation, une perspective empirique

1. Méthode
2. Résultats
3. Conclusions

I.B. Alkin, Daillak, Withe (1979) : l'évaluation fait-elle une différence ?

1. Définition : qu'est ce que l'utilisation ?
2. Méthode
3. Résultats
 - a) Le contexte préexistant à l'évaluation
 - b) Les orientations des utilisateurs
 - c) Les approches de l'évaluateur
 - d) La crédibilité de l'évaluateur
 - e) Les facteurs organisationnels et institutionnels internes
 - f) Les facteurs institutionnels externes
 - g) Le contenu de l'information et le rapport
 - h) Le style de l'administrateur
4. Conclusions

I.C. Une revue de littérature

1. Les définitions de l'utilisation
 - a) Sur le plan instrumental
 - b) Sur le plan conceptuel
 - c) Le plan de la prise d'information
 - d) Le plan "potentiel"
2. Méthodologie
 - a) Selon Cousins et Leithwood, on trouve trois types de recherche :
rétrospective, longitudinale et expérimentale
La recherche rétrospective
La recherche longitudinale
La recherche expérimentale
 - b) Leviton et Hughes comme Alkin et Daillak dressent un tableau très mitigé des résultats des recherches sur l'utilisation de l'évaluation
3. Synthèse des résultats
4. Conclusions

II. LA PRATIQUE DE L'EVALUATION

II.A. L'étude des théories

1. Alkin et Ellett (1979) : une analyse naturaliste
2. Williams (1989) : une analyse numérique
 - a) Méthodologie
Phase 1 : la mesure
Phase 2 : la classification
 - b) Résultats

Approche théorique

Approche pratique

Comparaison des approches pratique et théorique

c) Discussion

II.B. L'étude des pratiques d'évaluation

1. Une description des pratiques d'évaluation des politiques de la commission européenne

a) Définition de l'évaluation

La dimension affective

La dimension historique

La dimension politico-stratégique

La dimension pratique

b) Méthode

c) Résultats

d) Conclusions

2. Une enquête sur les pratiques en regard des "Standards"

3. Une enquête sur les pratiques des évaluateurs américains

a) Résultats

Profil professionnel

Evolution des pratiques

Contraintes

Influences théoriques

Pratiques d'évaluation

Une analyse factorielle a permis de dégager quatre types de pratiques indépendantes. Ces quatre types expliquent 45 % de la variance

Analyse prédictive

b) Synthèse

c) Conclusions

4. Conclusions

a) L'objet d'étude

b) La méthodologie

CHAPITRE 4

L'EVALUATION DE L'EFFICACITE ET DE LA QUALITE EN SANTE PUBLIQUE ET EN PROMOTION DE LA SANTE

I. INTRODUCTION : L'ANALYSE DE QUELQUES CAS

II. L'EFFICACITE ET LA QUALITE EN SANTE PUBLIQUE

II.A. Le transfert de la connaissance vers l'action

1. La décision dans la pratique médicale et en santé publique

a) La relation entre savoir éprouvé et pratique médicale

b) Les relations entre l'information sanitaire et la décision en santé publique

c) Des mécanismes pour assurer ce transfert entre connaissance et action

2. L'analyse d'une conférence de consensus

a) La présentation et l'analyse de la conférence de consensus par Broclain (1994)

b) Les propositions de Broclain à la lueur de la grille d'analyse APEP

II.B. La mesure de l'efficacité en santé publique et en santé au travail

1. Le changement de paradigme du processus d'évaluation en santé publique
2. L'évolution de la mesure des résultats dans le domaine de la santé au travail
3. La qualité méthodologique des études d'intervention dans le domaine de la santé au travail

II.C. L'accréditation et l'assurance de qualité

II.D. Synthèse sur l'approche de l'efficacité et de la qualité en santé publique

III. L'EFFICACITE ET LA QUALITE DE LA PROMOTION DE LA SANTE ET DE L'EDUCATION POUR LA SANTE

III.A. Pourquoi poser la question de l'efficacité et de la qualité ?

1. L'efficacité clinique
2. L'efficacité économique et politique
3. L'enjeu professionnel

III.B. Comment résoudre la question de l'efficacité ?

1. Réduire le "fossé"
2. Un outil d'analyse des études d'efficacité de la promotion de la santé et de l'éducation pour la santé
 - a) Le projet consiste à collecter des informations sur l'efficacité et sur la qualité de la recherche en évaluation et à la rendre plus accessible par une base de données et une série d'articles d'analyse
 - b) Veen et al. analysent le résultat de la démarche en s'appuyant sur la manière de remplir les grilles d'analyse, sur le contenu de ces grilles et sur les commentaires des analystes
3. Tenir compte de la spécificité
 - a) Le choix des objets de l'évaluation de l'efficacité
 - Les résultats épidémiologiques*
 - Les résultats comportementaux et cognitifs*
 - Les résultats attitudinaux, sociaux, environnementaux*
 - Les résultats sur les processus de "gestion" de la santé*
 - Les résultats sur la qualité de vie*
 - b) Les contraintes des méthodes d'intervention
 - La rigueur du schéma expérimental versus la pertinence et l'adaptabilité du programme*
 - La validité interne ou efficacité "vraie" versus la validité externe ou faisabilité*
 - Les effets expérimentaux versus les effets placebo*
 - L'efficience versus l'économie d'échelle*
 - Le bénéfice versus le coût*
 - Les effets à long terme versus les effets à court terme*
 - Le niveau d'investissement seuil versus les niveaux d'investissement à rendement décroissant des dépenses en éducation pour la santé*
 - L'application de ces dilemmes à la promotion de la santé*
4. Des remèdes
 - a) Les objets
 - b) Les méthodes

Se limiter à l'utilisation de données quantitatives
Développer les outils existants
Combiner les méthodes quantitatives et qualitatives
Développer de nouvelles approches

c) Le processus

Certaines pratiques mettent l'accent sur la composante jugement
D'autres pratiques permettent de "connaître" l'action avant de la juger
D'autres pratiques enfin sont directement au service de la décision

III.C. Comment résoudre la question de la qualité ?

1. Le développement du concept de la qualité dans les entreprises
2. Le développement du concept de la qualité dans les services de santé
3. L'équivalence entre l'évaluation de programmes ou de service et l'évaluation de qualité
4. Une proposition : distinguer évaluation de la qualité et assurance de la qualité
5. Une clarification terminologique
 - a) L'article de Deccache (1997) dénote une difficulté à percevoir la totalité de la démarche d'évaluation
 - b) Mac Donald (1997) propose de joindre les indicateurs de qualité et d'efficacité
6. L'évaluation de la qualité de la planification

IV. CONCLUSIONS : LA FECONDITE DU MODELE