

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE DESCRIPTIVE DES PRATIQUES ÉVALUATIVES :
RECOURS À LA LOGIQUE DE L'ÉVALUATION

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉDUCATION

PAR

JULIE DUVAL

JANVIER 2006

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je souhaite remercier ma directrice, Mme Marthe Hurteau, qui m'a soutenue tout au long de la production de cette recherche. Sa patience, son appui infaillible, sa disponibilité et sa très grande connaissance du sujet ont été pour moi source d'inspiration et d'encouragement. Je tiens aussi à souligner le soutien inestimable que m'a apporté Mme Carmen Parent, à titre de directrice de la maîtrise, tout au long de ma formation de 2^e cycle. Ses conseils ont toujours été avisés et justes.

Je remercie M. Sylvain Houle, professeur à l'École des sciences de la gestion de l'UQAM, et M. Jean Bélanger pour leur contribution à la validation de la grille d'analyse des articles. Je suis également reconnaissante aux membres de mon jury, M. Juan Wood et M. Pierre Toussaint, pour m'avoir lue et pour l'intérêt qu'ils ont manifesté par rapport à cette recherche.

Enfin, je désire exprimer ma reconnaissance envers ma famille, qui m'a appuyée dans cette démarche de mille façons, rendant ainsi possible l'aboutissement de cette recherche et de ma formation dans les temps voulus. Sans oublier mes très chères amies Marlène et Yousr, pour leur contribution morale et intellectuelle. Merci à tous d'avoir rendu possible la réalisation de ce mémoire.

TABLE DES MATIÈRES

Liste des figures	vii
Listes des tableaux	viii
Liste des abréviations	ix
RÉSUMÉ	x
INTRODUCTION	1
CHAPITRE I : PROBLÉMATIQUE	2
1. Le contexte	2
2. L'évolution de la discipline.....	3
3. Le problème	5
4. La question de la recherche.....	5
5. La pertinence de l'étude	6
CHAPITRE II : CADRE THÉORIQUE	8
1. L'évaluation de programme	8
2. La logique de l'évaluation.....	10
2.1 Les choix stratégiques et méthodologiques de Hurteau (1991).....	12
2.2 Les deux niveaux de logique de Fournier (1995).....	13
2.3 Les deux types de pensée de Stake (2004)	15
2.4 La synthèse des approches	15

3. Les paramètres de la logique de l'évaluation.....	17
3.1 La problématique et la question d'évaluation.....	17
3.2 Les objectifs et les types d'évaluation.....	18
3.3 L'objet et les critères.....	19
3.4 Les variables observables.....	20
3.5 Les standards.....	20
3.6 Le jugement.....	21
4. Les autres variables.....	22
4.1 Les cadres théoriques de référence.....	22
4.2 La contribution des modèles d'évaluation.....	23
4.2.1 Les distinctions entre la logique de l'évaluation et les modèles d'évaluation de programme.....	24
5. Les objectifs de la recherche.....	27
CHAPITRE III : MÉTHODOLOGIE.....	28
1. Le type de recherche.....	28
2. Les choix méthodologiques.....	28
2.1 Le choix de l'échantillon.....	29
2.2 La collecte des données.....	30
3. L'analyse et l'interprétation des données.....	31
4. Les limites de la recherche.....	32

CHAPITRE IV : RÉSULTATS	34
1. Objectif 1 : Le recours aux paramètres de la logique de l'évaluation.....	36
1.1 Les résultats descriptifs	36
1.1.1 La description du programme.....	38
1.1.2 La problématique.....	39
1.1.3 La ou les questions	40
1.1.4 L'objectif d'évaluation et/ou le type d'évaluation	41
1.1.5 L'objet d'évaluation	42
1.1.6 Les critères	42
1.1.7 Les variables observables	43
1.1.8 Les standards	44
1.1.9 Le jugement.....	45
1.2 L'analyse des cas.....	46
2. Objectif 2 : L'influence d'autres variables.....	54
2.1 La présence d'un cadre théorique de référence	55
2.2 Le recours à un modèle d'évaluation.....	56
3. Objectif 3 : L'évolution dans les pratiques	57
4. Objectif 4 : Le fondement du jugement	58
5. La synthèse des résultats	59
 CHAPITRE V : DISCUSSION	 60
 CONCLUSION	 63

RÉFÉRENCES.....	65
APPENDICE A : GRILLE D'ANALYSE.....	69
APPENDICE B : COEFFICIENT D'AGRÉMENT INTERJUGE	76
APPENDICE C : COMPILATION DES RÉSULTATS	77

LISTE DES FIGURES

Figure 2.1. Les choix stratégiques et méthodologiques (Hurteau, 1991)	12
Figure 2.2. Les deux niveaux de logique de Fournier (1995)	14
Figure 2.3. La synthèse des paramètres de la logique de l'évaluation	16

LISTE DES TABLEAUX

Tableau 2.1. Typologie de Rossi, Lipsey et Freeman (2004)	19
Tableau 2.2. Classification des modèles d'Alkin et Christie	23
Tableau 2.3. Distinction entre les modèles d'évaluation et la logique de l'évaluation	25
Tableau 4.1. Recension des articles analysés	34
Tableau 4.2. Distribution des fréquences pour chaque paramètre	37
Tableau 4.3. Résultats par article	47
Tableau 4.4. Profil d'évaluations	51
Tableau 4.5. Tableau croisé entre le type d'évaluation et les standards	53
Tableau 4.6. Résultats pour les variables périphériques	55

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

UQAM : Université du Québec à Montréal

RÉSUMÉ

L'évaluation de programme se voit remise en question régulièrement au cours des trente dernières années quant à sa capacité à produire un jugement de valeur utile à la prise de décision, et ce, malgré le développement important de cadres conceptuels et de méthodes. Guba (1972) et Scriven (1995) ont émis l'hypothèse que ces critiques pouvaient être imputables au faible recours à la théorie par les praticiens. De ce fait, il devient difficile pour ces derniers d'identifier des critères pertinents et de les appliquer adéquatement. Ce constat soulève la question des assises théoriques qui fondent la démarche évaluative. En 1980, Scriven a proposé un raisonnement fondamental, la logique de l'évaluation, susceptible de structurer cette dernière (Scriven, 1980). D'autres auteurs (Hurteau, 1991; Fournier, 1995; Stake, 2004) ont repris le concept de la logique de l'évaluation en proposant des approches différentes pouvant possiblement compléter la vision de Scriven (1980).

L'hypothèse proposée par Guba (1972) et Scriven (1995) n'ayant pas fait l'objet d'une étude, dont les résultats auraient pu la confirmer ou l'infirmer, la présente recherche vise précisément à décrire dans quelle mesure les praticiens en évaluation de programme réfèrent à la logique de l'évaluation pour structurer leur démarche évaluative. Pour ce faire, 69 articles tirés de trois revues spécialisées dans le domaine ont été analysés au moyen d'une grille descriptive.

Les résultats de ce mémoire semblent démontrer que les praticiens présentent leur démarche en ayant effectivement recours aux composantes de la logique de l'évaluation. Toutefois, il est impossible de déterminer si celle-ci a servi de cadre de référence dans la planification et la réalisation de la démarche évaluative, ou si elle a seulement servi de cadre pour la rédaction à des fins de publication. Dans tous les cas, il ressort que cette structure peut être considérée comme une pratique exemplaire et, en soi, cela demeure un constat susceptible d'alimenter de futures recherches quant aux aspects problématiques de la pratique évaluative.

Mots clés : Évaluation de programme, logique de l'évaluation, modèles d'évaluation, critères, standards, jugement.

INTRODUCTION

La présente recherche a pour objet d'étude les pratiques en évaluation de programme. Cette discipline a connu, au cours des quarante dernières années, un essor important qui se traduit actuellement par une littérature abondante. Cependant, cela ne l'empêche pas d'être confrontée régulièrement à des critiques quant à sa capacité à répondre aux attentes des clients et à produire de l'information utile à la prise de décision. Dès les années 1970, des auteurs importants (Guba, 1972; Scriven, 1980, 1995) ont soulevé cette difficulté rencontrée dans la pratique évaluative et ont identifié des pistes d'explication. Ils concluent, principalement, que les chercheurs comme les praticiens recourent peu aux fondements théoriques établis dans la littérature pour effectuer leurs évaluations et appuyer leurs jugements, ce qui affecte la validité de ces dernières. Christie (2003) confirme d'ailleurs cette hypothèse dans sa recherche sur la mise en application des modèles d'évaluation de programme dans la pratique évaluative, en concluant que les praticiens déclarent effectivement y avoir recours dans une très faible proportion. Ce constat amène l'auteur à se questionner sur les alternatives qu'ils utilisent dans la pratique pour structurer leurs démarches évaluatives.

L'hypothèse soulevée par Guba (1972) et Scriven (1995), à l'effet que la validité des jugements évaluatifs est remise en question à cause du manque de fondement théorique de la démarche, n'a jamais été vérifiée. La présente étude vise donc à décrire cette hypothèse en ce qui a trait à l'utilisation de la logique de l'évaluation comme fondement théorique de la démarche. La méthodologie de l'étude se fonde sur une analyse documentaire d'articles tirés de revues scientifiques, qui a été réalisée en fonction de paramètres clairement identifiés et définis dans le cadre théorique. L'analyse de l'information recueillie a été effectuée par la suite afin de dégager des éléments de réponse à la question soulevée.

Enfin, les chapitres suivants élaboreront l'ensemble des éléments nécessaires à la compréhension et à la réalisation de cette recherche, soit la problématique, le cadre théorique et la méthodologie, l'analyse des résultats et la discussion.

CHAPITRE I

PROBLÉMATIQUE

1. LE CONTEXTE

Selon Fournier (1995) le développement d'un raisonnement évaluatif conduisant à un jugement légitimé quant au programme à l'étude pourrait bien constituer la pierre angulaire d'un échange réussi entre la théorie et la pratique évaluative. Dans cette perspective, la récente étude de Christie (2003) tend à démontrer que les praticiens ont peu recours aux modèles d'évaluation mis à leur disposition, ce qui amène l'auteur à conclure que « one of the next obvious question is, if formal theory does not guide their work, then what does? Research on evaluation can help answer this question.» (Christie, 2003, p. 92.) Or, considérant les importants développements théoriques et méthodologiques qu'a connus la discipline évaluative au cours des trente dernières années, il est surprenant de constater que peu d'études ont été réalisées sur la pratique évaluative, et encore moins sur ce qui pourrait constituer un pont entre cette théorie et la pratique. Notre recherche s'inscrit dans cette perspective en tentant de fournir des éléments de réponse en regard de l'hypothèse émise par Guba (1972) et Scriven (1995) pour expliquer l'échec de la fonction d'évaluation de programme, soit par l'étude de la logique de l'évaluation comme pont entre la théorie et la pratique.

2. L'ÉVOLUTION DE LA DISCIPLINE

Avant d'aborder le sujet même de cette recherche, soit la logique de l'évaluation, il nous apparaît pertinent de la situer dans le contexte de l'évolution de la discipline.

L'évaluation est souvent considérée comme un phénomène récent, alors que son origine remonte à plus de 150 ans. Cette perception s'explique en grande partie par le fait que l'évaluation n'a connu un essor important qu'à la fin des années 1960, période au cours de laquelle les chercheurs en sciences sociales ont été mis à contribution afin d'évaluer l'efficacité des nombreux programmes de grande envergure mis de l'avant à la fin de la Seconde Guerre Mondiale (Madaus et Stufflebeam, 2004). La recherche évaluative revêtait alors deux fonctions, à savoir celle de juger de l'efficacité des programmes et celle d'établir le rendement des organisations dépendantes des fonds publics. À cette époque, l'évaluation était étroitement associée à la recherche évaluative, lorsque les deux activités n'étaient pas carrément confondues. D'ailleurs, cette confusion semble encore persister de nos jours (Hurteau et Houle, 2005).

L'évaluation s'est vraiment distinguée de la recherche évaluative pour devenir une discipline en soi au début des années 1980, lorsqu'il est devenu évident que les besoins et les intérêts des clients différaient sensiblement de ceux des chercheurs. L'évaluation s'est alors orientée davantage vers la pratique et la prise de décision que vers le développement des connaissances, afin de tenter de mieux répondre aux besoins d'information des clients (Stake, 2004). La période de professionnalisation qu'a connue l'évaluation au cours des années 1970 et 1980 a donné lieu à l'élaboration d'une multitude de modèles et de méthodes visant à structurer la démarche d'évaluation (Madaus et Stufflebeam, 2004). Dans la foulée de ces développements, Scriven (1980) a introduit le concept de logique de l'évaluation qui constitue, selon cet auteur, le fondement même de toute évaluation. Cette logique réfère à la capacité

de l'évaluateur de poser le « bon » jugement en utilisant le « bon » raisonnement et comporte les quatre opérations fondamentales suivantes :

1. Déterminer les critères;
2. Établir les standards;
3. Mesurer la performance et la comparer aux standards établis;
4. Analyser et intégrer les résultats dans un jugement sur le mérite ou la valeur de l'objet d'étude (Scriven, 1980).

Le concept de la logique de l'évaluation a été repris et développé par d'autres auteurs, tels que Hurteau (1991), qui a introduit la notion de choix stratégiques, Fournier (1995), qui a intégré la notion de logique de l'évaluation dans son concept de logique générale, et, plus récemment, Stake (2004), qui a proposé une pensée basée sur les critères. Leur contribution a été significative principalement sur deux plans. D'une part, ils ont situé ces quatre opérations au sein de l'ensemble de la démarche évaluative. D'autre part, ils ont dégagé l'ensemble des paramètres requis afin de structurer la démarche évaluative en un tout cohérent, à savoir : la problématique, la question, l'objet, les critères, les variables observables, les standards et le jugement posé. L'ensemble de ces paramètres sera repris et développé dans le cadre théorique.

Finalement, il est à noter que la vague d'expansion des années 1980 s'est poursuivie dans les années 1990. Cette période se caractérise principalement par l'émergence de modèles orientés vers les clients, tel le modèle participatif (« empowerment evaluation »). Les années 2000 montrent une volonté de consolider ces acquis sous la forme de formations accréditées, d'activités de développement professionnel offertes par les associations professionnelles et de recherche sur l'évaluation. Alors que la recherche sur l'évaluation des années 1990 s'était principalement concentrée sur l'identification des problèmes rencontrés dans la pratique évaluative, la recherche des années 2000 commence à se pencher sur leurs causes et leurs solutions.

3. LE PROBLÈME

L'évolution de la discipline illustre que, malgré un essor important et un développement conceptuel et méthodologique non moins négligeable, l'évaluation de programme n'arrive toujours pas à répondre aux attentes de ses clients en produisant de l'information utile à la prise de décision en temps opportun (McDavid, 2001). Certains auteurs (Guba, 1972; Scriven, 1995) ont émis l'hypothèse que cette remise en question résulte d'un manque de validité du jugement évaluatif en raison de fondements théoriques insuffisants. Cette hypothèse n'a cependant jamais fait l'objet d'une vérification dans la pratique.

Dans la foulée de ce qui a été avancé par Fournier (1995) quant à la nécessité de mettre en relation la théorie et la pratique évaluative, et dans la foulée des travaux de Christie (2003), qui conclut que les modèles théoriques d'évaluation ne sont pas utilisés par les praticiens, il apparaît de plus en plus évident qu'il faut documenter la pratique afin de comprendre les problèmes qui en émanent et, ultimement, envisager des solutions. La logique de l'évaluation proposée par Scriven (1980) se veut un pont entre la théorie et la pratique en constituant un raisonnement fondamental susceptible de structurer toute démarche évaluative.

4. LA QUESTION DE LA RECHERCHE

Selon l'hypothèse de Guba (1972) et Scriven (1995), les remises en question connues par la pratique évaluative proviennent d'un manque de validité des conclusions en raison de leur manque d'appui théorique. Selon la définition du concept développé par Scriven (1980), la logique de l'évaluation pourrait bien constituer un fondement théorique susceptible de soutenir la démarche évaluative et d'en garantir la validité. Il devient cependant nécessaire, dans un premier temps, de déterminer dans quelle

mesure la logique de l'évaluation constitue une pratique exemplaire. La présente étude s'attache à cette préoccupation. Précisons qu'elle s'inscrit dans le cadre d'un programme de recherche dont elle constitue le premier volet. Dans cette perspective, elle vise à répondre à la question suivante

Dans quelle mesure les chercheurs et les praticiens de l'évaluation de programme réfèrent-ils aux paramètres de la logique de l'évaluation pour structurer leur démarche évaluative?

5. LA PERTINENCE DE L'ÉTUDE

L'évaluation de programme se conclut par un jugement sur la valeur en vue de prendre une décision juste et éclairée. Étant donné les conséquences associées à cette prise de décision, il importe de se questionner sur la démarche qui conduit un évaluateur à poser un tel jugement. Ainsi, la pertinence de cette étude est de deux ordres : scientifique et social.

En ce qui concerne la pertinence scientifique, l'évaluation est remise en question quant à sa capacité à produire des conclusions valides. Dubois et Marceau (2005) soulignent que la discipline de l'évaluation de programme stagne dans une révolution scientifique parce qu'elle ne se fonde pas sur un corpus théorique, méthodologique et empirique harmonisé. Pour que l'évaluation passe au quatrième stade de l'évolution d'une science, c'est-à-dire le passage à une nouvelle science normale, il faut que la théorie traditionnelle post-positiviste s'harmonise avec le nouveau courant théorique constructiviste. Un des facteurs pouvant expliquer la faiblesse des échanges fructueux entre ces deux noyaux théoriques proposés par Dubois et Marceau (2005) est le manque de solidité dans l'arrimage entre les deux niveaux de connaissances (théoriques et pratiques). Ce mémoire s'inscrit dans ce mouvement de recherche de scientificité en évaluation de programme puisqu'il vise à établir dans quelle mesure la logique de l'évaluation, en tant que cadre théorique, est utilisée dans la pratique.

L'évaluation de programme est au cœur des systèmes éducatifs nord-américains parce qu'elle assure une rétroaction aux différentes parties prenantes et qu'elle éclaire la prise de décision. L'ampleur des conséquences liées à cette décision exige une grande rigueur. La logique de l'évaluation propose donc une démarche rigoureuse permettant de structurer toute démarche évaluative et de produire ainsi des conclusions valides respectant une certaine justice sociale. Ainsi, ces constats confèrent à l'évaluation de programme une pertinence sociale indéniable.

En conclusion, la logique de l'évaluation, en permettant de produire des conclusions valides, permet aux décisions qui en découlent d'être plus justes socialement parce que fondées scientifiquement.

CHAPITRE II

CADRE THÉORIQUE

La problématique soulevée dans le chapitre précédent fait référence à la mise en application des paramètres de la logique de l'évaluation. Dans cette perspective, le présent chapitre abordera les aspects théoriques nécessaires à la mise en opération de cette problématique, soit la définition de l'évaluation de programme, la définition de la logique de l'évaluation, de même que sa distinction des modèles d'évaluation, les paramètres qui composent la logique de l'évaluation et les autres variables périphériques à l'étude, soit les cadres théoriques de référence et les modèles d'évaluation.

1. LA NOTION DE PROGRAMME

L'évaluation de programme a pour objet d'évaluation un programme. Hurteau et Houle (2005) définissent un programme comme un système ouvert et systémique composé de différentes dimensions, soit les intrants, le processus, les résultats, l'environnement et la rétroaction. S'inspirant des travaux de Chen (2005), ces auteurs définissent les composantes d'un programme comme suit. Les intrants se composent des clients du programme, de leurs besoins ainsi que des ressources de l'environnement du programme. Le processus, quant à lui, transforme les intrants en résultats et prend la forme d'interventions auprès de la clientèle. Ces dernières produisent des changements visant à répondre aux besoins de la clientèle ayant motivés la mise sur pied du programme. Finalement, ces changements bénéficient à l'environnement qui voit là l'aboutissement de l'utilisation de ses ressources par la

satisfaction des besoins de sa clientèle. La rétroaction est un retour d'information qui permet au programme de s'ajuster et d'éviter de se détériorer. La rétroaction est souvent générée par la démarche évaluative qui vise à produire un jugement de valeur devant servir à prendre des décisions quant à la survie ou l'amélioration du programme.

En somme, un programme est un ensemble de ressources organisées en vue de produire un changement planifié par le biais d'intervention ciblée. En ce sens, il s'adresse à une clientèle précise, la nature du changement attendu est établi en termes observables, mesurables et appréciables et les modalités et les délais pour y parvenir sont précisés. Enfin, le changement s'inscrit dans un contexte environnemental distinct.

2. L'ÉVALUATION DE PROGRAMME

Rossi, Lipsey et Freeman (2004), définissent l'évaluation de programme comme un processus systématique de collecte de donnée visant à poser un jugement sur la valeur d'un objet. Cette définition est reprise par d'autres auteurs tels que Wholey, Hatry et Newcomer (2004) et le Joint Committee (1994). Stake (2004) spécifie que le jugement devrait se baser sur des standards qui peuvent être soit externes ou absolus, soit relatifs ou comparatifs. Ces différentes définitions intègrent plusieurs éléments fondamentaux d'une démarche évaluative, à savoir : l'objet, les critères et les indicateurs (ou les variables observables), les standards et le jugement.

Tout récemment Hurteau et Houle (2005) ont proposé une nouvelle perspective en définissant l'évaluation de programme comme une démarche qui consiste à porter un jugement de valeur sur un programme en vue de prendre des décisions sur sa

programmation future. Cette démarche implique une cueillette systématique de données portant sur les composantes du programme – intrants, processus et résultat – et s’inscrit dans une perspective de gestion des programmes. Cette définition adopte la position de plusieurs théoriciens de l’évaluation, tel Patton in Alkin (1990); Scriven (1991); Patton (2002) et Stake (2004).

En résumé, l’évaluation de programme s’avère une démarche rigoureuse qui vise à produire un jugement de valeur sur un objet d’évaluation, le programme, ou encore une ou plusieurs de ces composantes, en vue de produire de l’information utile à la prise de décision.

2. LA LOGIQUE DE L’ÉVALUATION

Scriven (1995) mentionne que l’évaluation est une forme d’enquête et que, de ce fait, elle a besoin d’une base légitime pour supporter ses conclusions, sans quoi la validité de ces dernières est remise en question. Ainsi, comme le mentionnent Toulmin (1964, 1984), McCarthy (1973, 1979), Redding (1989) et Taylor (1961), chaque forme d’enquête possède un patron général de raisonnement transdisciplinaire qui guide et informe la pratique. Ce « patron général » est un raisonnement fondamental, une sorte de supra-structure, qui permet de légitimer les conclusions. La logique de l’évaluation suggérée par Scriven (1980) propose les balises de ce raisonnement fondamental, cette supra-structure propre à l’évaluation de programme qui la distingue, par le fait même, des autres types d’enquête.

Le concept de la logique de l'évaluation a d'abord été défini par Scriven en 1980. À la base, il comprend les quatre étapes suivantes, déjà mentionnées dans la problématique :

1. Déterminer les critères : par rapport à quels éléments ou composantes l'objet d'étude doit-il performer?
2. Établir les standards : quel est le niveau de performance attendu?
3. Mesurer la performance et la comparer aux standards établis : dans quelles mesures l'objet a-t-il performé, répondu aux standards?
4. Analyser et intégrer les résultats dans un jugement sur le mérite ou la valeur de l'objet d'étude : quelle est la valeur de l'objet d'étude? (Scriven, 1980.)

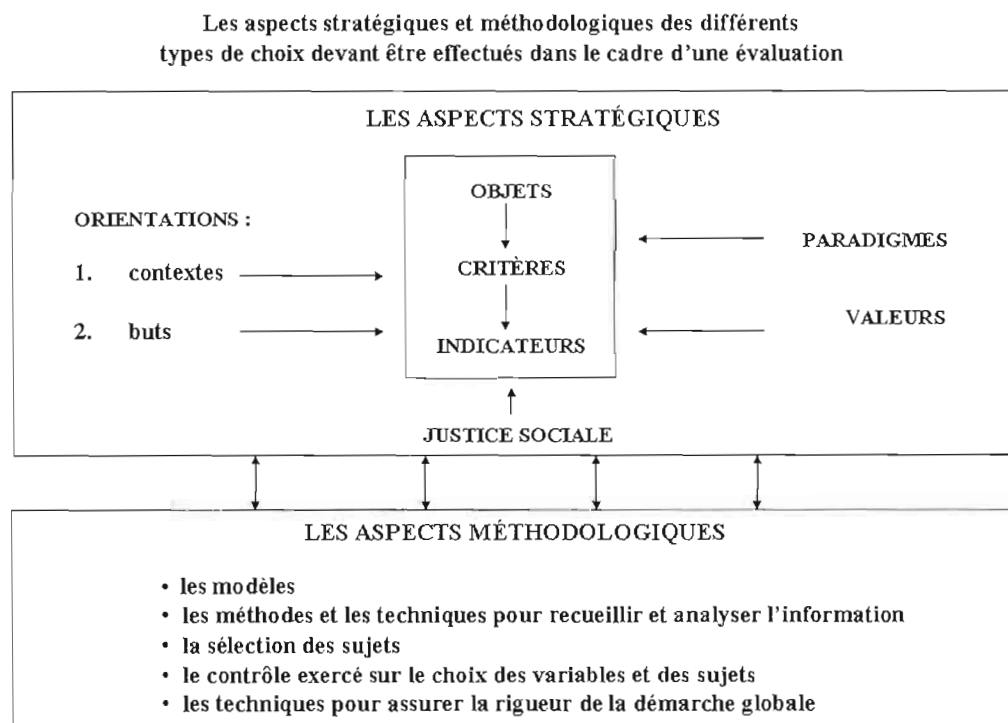
Ces étapes contiennent les paramètres fondamentaux de la logique de l'évaluation, soit les critères, les standards, la mesure ou les variables observables et le jugement. Ces paramètres transcendent les conditions propres à chaque évaluation et devraient, par conséquent, représenter la base de toute démarche évaluative afin de garantir la validité des conclusions.

Tout en reconnaissant la valeur fondamentale des quatre opérations présentées par Scriven en 1980, il nous apparaît évident qu'elles doivent s'inscrire dans une structure d'ensemble afin d'être utilisables dans le cadre d'une démarche évaluative. Certains auteurs (Hurteau, 1991; Fournier, 1995; Stake, 2004), mentionnés aussi dans la problématique, ont poursuivi la réflexion amorcée par Scriven (1980) afin, d'une part, de dégager des paramètres permettant d'intégrer ces opérations dans le cadre opérationnel de la pratique et, d'autre part, de tenir compte des développements conceptuels et méthodologiques survenus depuis la définition proposée par Scriven. Les sections suivantes présentent ces approches de façon plus détaillée.

2.1 Les choix stratégiques et méthodologiques de Hurteau (1991)

Hurteau (1991) distingue deux types de choix au sein d'une même évaluation : les choix stratégiques et les choix méthodologiques. Les choix stratégiques comportent l'identification et la définition de l'objet d'étude, de ses critères et des indicateurs qui serviront à le mesurer en fonction de considérations telles que l'orientation de la démarche (ses buts et son contexte) les valeurs, les paradigmes en cause dans l'évaluation et la notion de justice sociale. Les choix méthodologiques, quant à eux, découlent des choix stratégiques et précisent le déroulement de l'évaluation sur le plan de la cueillette et de l'analyse de l'information.

Figure 2.1. Les choix stratégiques et méthodologiques (Hurteau, 1991)



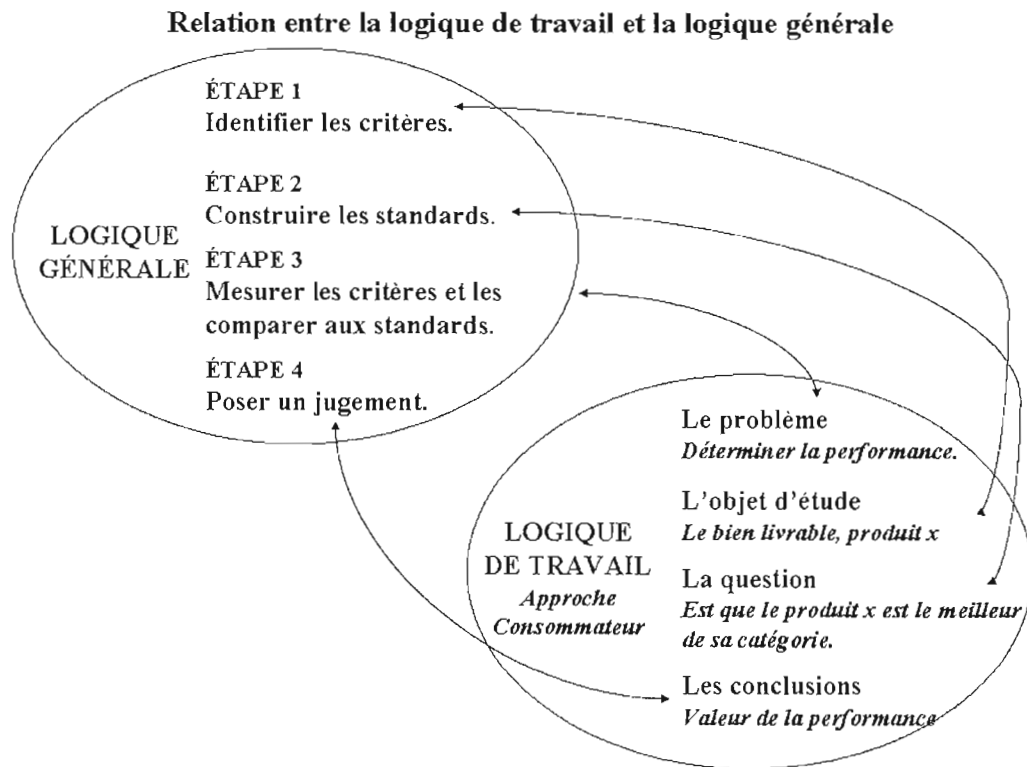
(Hurteau, 1991 ; p-4)

Selon Hurteau (1991), les choix stratégiques doivent être énoncés au départ de toute démarche évaluative. En effet, les choix stratégiques servant à encadrer les choix méthodologiques, il convient de les effectuer en premier afin de limiter les risques de dérive et de faire en sorte que le jugement soit valide et documenté et qu'il produise des recommandations éclairantes. Ainsi, la distinction faite par Hurteau (1991) entre les deux types de choix évaluatifs met l'accent sur l'importance des choix stratégiques comme ancrage de toute démarche évaluative qui se veut logique.

2.2 Les deux niveaux de logique de Fournier (1995)

Fournier (1995) poursuit l'idée de la logique d'évaluation de Scriven (1980) en proposant deux niveaux de logique. Le premier, soit la logique générale, intègre textuellement les étapes de la logique de Scriven (1980). Ce niveau spécifie les règles du jeu et indique comment l'évaluation devrait être réalisée, indépendamment du champ d'expertise ou du contexte. Ce niveau est crucial puisqu'il permet d'identifier les paramètres fondamentaux à la réalisation de toute évaluation qui prétend produire un jugement éclairant. Le second niveau, la logique de travail, est plus opérationnel. Cette logique reprend les éléments du premier niveau en fonction du contexte propre à chaque évaluation. Ainsi, dans la pratique, il est possible d'observer que la logique générale se traduit par une multitude de logiques de travail. En somme, selon Fournier (1995), la logique générale fait appel à la dimension stratégique de la démarche, alors que la logique de travail réfère au volet tactique.

Figure 2.2. Les deux niveaux de logique de Fournier (1995)



(Fournier, 1995, p.20)

Cette approche propose une correspondance intéressante entre les paramètres de la logique générale, plus stratégiques, et ceux de la logique de travail, qui se veulent plus opérationnels. Il est aussi à noter que les flèches sont à double sens, illustrant encore une fois l'importance de l'influence réciproque entre le stratégique et le méthodologique, entre les éléments conceptuels et les éléments opérationnels.

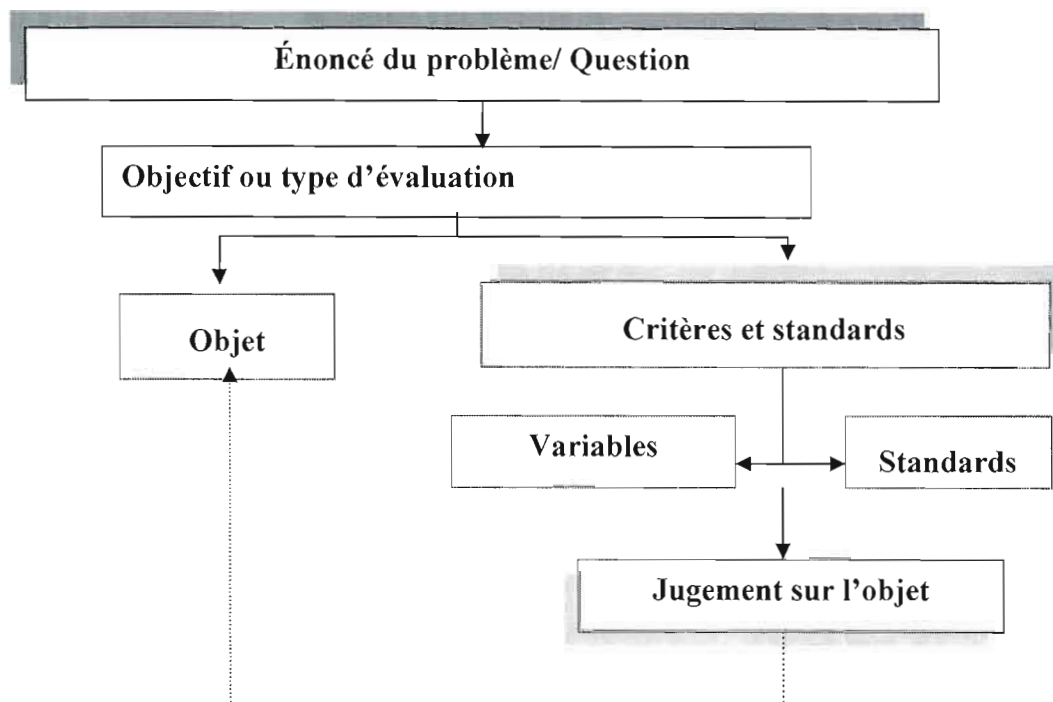
2.3 Les deux types de pensée de Stake (2004)

Stake (2004) propose deux types de pensée évaluative susceptibles de fonder des évaluations différentes. « I am calling the one kind of thinking criterial to build upon the analysis of descriptive variables and the other interpretive to build upon experiential, personal knowing in real space and real time and with real people. » (Stake, 2004, p. xv.) Stake (2004) mentionne ensuite que la pensée fondée sur les critères est utilisée dans des évaluations basées sur les standards (« standard-based »), alors que la pensée de nature interprétative se retrouve surtout dans des évaluations dites répondantes (« responsive »). L'évaluation basée sur les standards répond à une pensée rigoureuse fondée sur des éléments spécifiques à établir et s'appuie sur des mesures de performance, alors que l'évaluation répondante s'avère être davantage une attitude qu'une recette à suivre.

2.4 La synthèse des approches

La réflexion de ces différents auteurs souligne l'importance de structurer la démarche d'évaluation de programme afin de lui assurer une cohérence interne. Celle-ci est le résultat de la capacité de l'évaluateur à faire sienne la logique de l'évaluation, de saisir, comprendre et utiliser adéquatement les niveaux de choix, les types d'évaluation et les concepts qui sont au cœur de cette dernière. La figure suivante présente une synthèse des éléments retenus de cette démarche de réflexion selon une séquence applicable dans la pratique.

Figure 2.3. La synthèse des paramètres de la logique de l'évaluation



(Hurteau, 2005)

Ce graphique intègre plusieurs éléments proposés par Stake (2004) par rapport à l'évaluation basée sur les standards et les éléments des deux niveaux de logique tels que définis par Fournier (1995). Par ailleurs, les choix stratégiques présentés par Hurteau (1991) se retrouvent au début de la démarche afin de guider l'établissement des choix méthodologiques et des éléments subséquents de la séquence. Enfin, la synthèse présentée à la figure 2.3 offre un cadre d'analyse des démarches d'évaluation permettant d'établir leur niveau d'adéquation avec la logique de l'évaluation.

3. LES PARAMÈTRES DE LA LOGIQUE DE L'ÉVALUATION

La logique de l'évaluation, en tant que structure fondamentale, se compose de paramètres stratégiques et méthodologiques rigoureux susceptibles de garantir la validité des conclusions. Ainsi, la figure 2.3 (p. 17) présente les paramètres jugés essentiels dans les approches de Hurteau (1991), Fournier (1995) et Stake (2004) présentées précédemment. Les sections suivantes définissent chacun de ces paramètres plus en détail.

3.1 La problématique et la question d'évaluation

Point de départ de l'évaluation, la problématique influence, selon Fournier (1995) l'ensemble des paramètres puisqu'elle dicte l'orientation générale de l'évaluation. En ce sens, ce paramètre énonce les problèmes rencontrés par le responsable du programme qui ont justifié la réalisation de l'évaluation. Étant donné que les conclusions produites par l'évaluation ont pour objectif de répondre à cette problématique, il importe de bien identifier la problématique et de cibler les aspects à prioriser.

Ainsi, la problématique implique d'identifier les besoins du client et de pouvoir en tenir compte dans la formulation de la question d'évaluation. Cette dernière se veut une opérationnalisation de la problématique et doit donc être très étroitement en lien avec cette dernière.

3.2 Les objectifs et les types d'évaluation

Rossi, Lipsey et Freeman (2004) ont proposé une typologie de l'évaluation qui met en relation le type d'évaluation et les objectifs poursuivis. Cette typologie a été retenue pour sa rigueur et sa pertinence en matière d'évaluation et parce qu'elle constitue un point de référence reconnue dans le domaine de l'évaluation.

Cette classification comporte cinq types d'évaluation ayant une dépendance logique avec des objectifs qui leur sont spécifiques. Tout d'abord, l'évaluation de pertinence vise à démontrer la pertinence sociale du programme. L'évaluation de la conceptualisation et de la planification a pour objectif d'établir dans quelle mesure les services offerts répondent aux besoins de la clientèle. L'évaluation de l'implantation a pour but de déterminer dans quelle mesure le programme a été implanté et exécuté tel que planifié. Ensuite, l'évaluation de l'impact cherche à vérifier le niveau d'atteinte des résultats, des effets et des impacts attendus du programme. Enfin, l'évaluation de l'efficacité a pour visée d'identifier et de mesurer la performance des résultats, des effets et des impacts par rapport aux coûts du programme et à ceux d'autres programmes similaires.

Chaque type d'évaluation sous-tend un objectif d'évaluation bien précis, une évaluation pouvant comporter plusieurs objectifs. Chaque objectif doit être en lien avec la question d'évaluation et avoir pour but d'y répondre.

3.3 L'objet et les critères

L'objet d'évaluation consiste en la dimension du programme sur laquelle l'évaluation doit porter un jugement, alors que les critères sont les points de référence sur lesquels l'évaluateur s'appuiera pour poser un jugement objectif. La typologie proposée par Rossi, Lipsey et Freeman (2004), mentionnée précédemment, permet d'illustrer la relation entre le type d'évaluation et l'objectif poursuivi par celle-ci. Le tableau 2.1 établit la relation entre le type d'évaluation, les objets et les critères.

Tableau 2.1. Les différents objets et critères en relation avec la typologie élaborée par Rossi, Lipsey et Freeman (2004)

Type d'évaluation	Objet	Critères
Évaluation de pertinence (Needs assessment)	Objectifs et visées du programme	Problèmes à résoudre Besoins de la clientèle
Évaluation de la conceptualisation et de la planification (Assessment of program theory)	Activités, stratégies et organisation des services offerts par le programme	Objectifs du programme
Évaluation de l'implantation (Assessment of program process)	Réalisation du programme	Organisation du programme
Évaluation de l'impact (Impact assessment)	Résultats, effets et impacts du programme	Objectifs, visées et problèmes à résoudre
Évaluation de l'efficacité ou du coût-bénéfice (Efficiency assessment)	Résultats, effets et impacts du programme	Ressources investies

Dans la mesure où une évaluation comporte plus d'un objectif, chaque objectif donnera généralement lieu à un objet de recherche distinct et à différents critères, les critères étant propre à chaque objet d'étude. L'objet d'étude doit être en lien avec l'objectif de l'évaluation (ou un des objectifs s'il y en a plusieurs) et les critères, quant à eux, doivent être en lien avec l'objet d'étude.

3.4 Les variables observables

Le paramètre des variables observables, aussi appelées indicateurs, permet de mesurer l'objet d'étude et d'obtenir les observations à partir desquelles les écarts seront calculés. À cause du développement et de l'importance grandissante de l'approche par indicateurs de performance utilisée dans le cadre d'une démarche de suivi des programmes, et pour éviter la confusion, nous avons décidé de retenir un terme qui est de plus en plus utilisé pour remplacer les indicateurs : celui de *variables observables*. Comme leur nom l'indique, il s'agit de variables à caractère observable dont la fonction est de mesurer les critères desquels elles découlent. Les variables observables se veulent une opérationnalisation des critères avec lesquels il convient d'observer un lien logique.

3.5 Les standards

Le terme *critère* renvoie souvent, dans la littérature anglaise, au terme de *standard*. Stake (2004) souligne qu'une confusion règne autour de ce concept. « The terms *criteria* and *standards* are not used in the same way by everyone, but most evaluators use *criterion* to mean an important descriptor or attribute and *standard* to mean the amount of that attribute needed for certain judgement. » (Stake, 2004, p. 5) Cette distinction sera retenue pour la présente étude puisqu'elle semble la plus couramment

utilisée. Scriven (1991), quant à lui, définit le standard comme le niveau de performance acceptable servant à déterminer si le critère a été atteint. Étant donné leur nature quantitative, il est possible que, dans certaines évaluations à caractère très qualitatif, l'usage des standards s'avère inadéquat. Dans la mesure où leur utilisation est appropriée, les standards doivent être en lien avec les critères ou les variables observables.

3.6 Le jugement

Fournier (1995) mentionne que le jugement représente la synthèse des données à la suite d'une pondération adéquate de leur importance. Le jugement repose sur l'analyse des écarts entre la mesure obtenue à partir des variables observables et les standards leur correspondant. Ainsi, il doit être en lien avec les critères, les variables observables et les standards. Un jugement est éclairant dans la mesure où il permet d'atteindre le ou les objectifs de l'évaluation et, par le fait même, de répondre à la question d'évaluation. Par ailleurs, il doit permettre de produire des conclusions d'évaluation valides, soit des conclusions fondées sur des mesures fidèles et valides. Scriven (1995) mentionne, à cet effet, que le jugement ne remplit son mandat qu'en respectant une démarche structurée et rigoureuse qui répond au besoin de cohérence entre les paramètres et qui garantit ainsi la crédibilité des conclusions qui en découlent. Dans le cas présent, la cohérence entre les paramètres se comprend par la correspondance entre la question, l'objectif de l'évaluation ainsi que le choix des objets, critères et standards.

4. LES AUTRES VARIABLES

Bien que cette recherche porte sur la logique de l'évaluation, il apparaît important de prendre en considération certaines variables qui sont susceptibles d'affecter l'utilisation des paramètres de la logique de l'évaluation, soit les cadres théoriques de référence et les modèles d'évaluation. Le choix d'étudier les modèles théoriques s'inspire d'une étude de Lipsey (1988) qui, quoique portant sur un sujet différent, met en évidence leur faible recours et les impacts significatifs de ce constat sur les conclusions des études. Lipsey (1988) a démontré dans son étude que les praticiens qui utilisent une méthodologie fondée sur les schémas expérimentaux et quasi-expérimentaux font peu appel à des cadres théoriques de référence pour établir le contexte théorique de leur évaluation devant servir à fonder leurs critères et leurs standards. Cette recherche tente de déterminer si le recours (ou l'absence de recours) à un cadre théorique de référence pour établir les critères de l'évaluation a un impact sur la présence (ou l'absence) de la logique de l'évaluation comme structure de l'évaluation dans les articles étudiés. Par ailleurs, Christie (2003) a établi dans une étude que les praticiens de l'évaluation utilisent peu les modèles d'évaluation développés précisément pour les aider à structurer leur démarche évaluative. Il est intéressant de vérifier dans quelle mesure le recours à des modèles d'évaluation peut être compatible avec l'utilisation de la logique de l'évaluation comme structure de l'évaluation, et dans quelle mesure il peut influencer cette utilisation.

4.1 Les cadres théoriques de référence

Tel que relevé dans l'étude de Lipsey (1988), les cadres théoriques de référence servent de base au choix des critères et des standards en se fondant sur une littérature pertinente à l'objet d'évaluation cadre de référence alors que les cadres athéoriques impliquent que les choix mentionnés sont effectués à partir de l'expérience de

professionnels, évaluateurs ou experts du terrain. Ce choix est soit non fondé ou est effectué en raison de la pauvreté de la littérature dans le domaine. La recherche de Lipsey (1988) l'a aussi amenée à conclure que le jugement des évaluations est affecté par le manque de validité du cadre théorique et de l'opérationnalisation de la démarche. Ces différentes considérations expliquent le choix actuel de s'intéresser aux modèles théoriques.

4.2 La contribution des modèles d'évaluation

La pratique de l'évaluation implique d'utiliser certains modèles d'évaluation qui guident la démarche en proposant un canevas devant orienter généralement les choix méthodologiques. Tel qu'illustré dans le tableau 2.2, Alkin et Christie (2004) proposent une classification des auteurs et de leurs modèles en fonction de la priorité qu'ils accordent à la méthodologie, au jugement ou à la prise de décision.

Tableau 2.2. Classification des modèles d'Alkin et Christie

Classification	Auteurs	
Modèles axés sur la méthode	<ul style="list-style-type: none"> • Campbell (1957) • Tyler (1942) • Schuman (1967) • Cook (1976) • Boruch (1977) 	<ul style="list-style-type: none"> • Rossi (1979) • Chen (1983) • Weiss (1972) • Cronbach (1977)
Modèles axés sur le jugement	<ul style="list-style-type: none"> • Scriven (1967, 1972) • Eisner (1976) • Owen (1973) • Wolf (1975) 	<ul style="list-style-type: none"> • Stake (1972) • House (1980) • Guba et Lincoln (1989)
Modèles axés sur la prise de décision	<ul style="list-style-type: none"> • Stufflebeam (1983) • Provus (1971) • Wholey (1981) • Patton (1978) • Alkin (1972) 	<ul style="list-style-type: none"> • Cousins (1992) • Preskill (1998) • King (1982) • Fetterman (1984)

Le choix d'un de ces modèles dans le cadre d'une démarche évaluative a nécessairement un impact sur l'ensemble des autres paramètres. Par exemple, le modèle de Schuman insiste sur les choix méthodologiques, tels que le devis et les outils, alors que celui de Fetterman accorde plus d'importance à la participation des personnes concernées, au détriment de la rigueur de la démarche. Dans le premier cas, les critères seront alors établis selon des principes issus de la méthode scientifique, alors qu'ils seront issus des choix des personnes concernées dans le second cas. En définitive, la liste des critères ainsi établis risque d'être plutôt différente.

On le rappelle, il importe de ne pas confondre ces modèles avec la logique de l'évaluation présentée précédemment. La section suivante expose les principales distinctions entre les modèles d'évaluation et la logique de l'évaluation.

4.2.1 Les distinctions entre la logique de l'évaluation et les modèles d'évaluation de programme

Il est facile de confondre la logique de l'évaluation et les modèles d'évaluation puisque tous les deux offrent, chacun à sa façon, un cadre de référence pour baliser la démarche évaluative dans la pratique. C'est pourquoi nous avons jugé important d'établir la distinction entre ces deux concepts.

Les modèles d'évaluation ont été développés depuis les années 1970 par nombre de praticiens de l'évaluation. Que l'on pense aux modèles plus traditionnels, avec les schémas expérimentaux et quasi expérimentaux (Campbell, 1957; Stanley, 1972; Cook et Campbell, 1976), ou aux plus récents, avec le modèle naturaliste de Guba et Lincoln (1989) et l'évaluation répondante de Stake (2004), tous ont pour objectif, selon Alkin et Christie (2004), d'orienter la démarche évaluative et de guider les choix méthodologiques. Ainsi, il y aura autant de démarches qu'il y a de modèles, et les choix effectués au cours de la démarche seront influencés par le paradigme sous-jacent au modèle utilisé.

Le tableau suivant résume les principales ressemblances et différences entre les modèles de l'évaluation et la logique de l'évaluation.

Tableau 2.3. Distinction entre les modèles d'évaluation et la logique de l'évaluation

	Modèles d'évaluation	Logique de l'évaluation
BUTS	Structurer la démarche évaluative en fonction de l'objectif de la démarche et du type de relation avec le client.	Offrir une supra-structure, un rationnel fondamental à toute démarche évaluative dans le but de générer un jugement adéquat.
PARADIGMES	Formalistes, naturalistes, etc.	
ORIENTATIONS	Méthode Prise de décision Jugement	Jugement
COMPOSANTES	Chaque modèle propose une méthodologie et des composantes adaptées aux objectifs.	Invariables : qui devraient se retrouver dans toute démarche évaluative.

La logique de l'évaluation, par son rationnel fondamental, transcende la structure proposée par le modèle en fonction de l'objectif et du type de relation avec le client. Les modèles se fondent sur un paradigme précis alors que la logique de l'évaluation se veut plus universelle. Dans un même ordre d'idées, un modèle peut être orienté vers la méthode, la prise de décision ou le jugement, selon la classification d'Alkin et

Christie (2004), tandis que la logique de l'évaluation s'oriente toujours vers la recherche d'un jugement valide et juste. Ainsi, il existe une multitude de modèles d'évaluation qui proposent autant de méthodologies et de composantes différentes qui s'adaptent aux objectifs de la démarche. Au contraire, la logique de l'évaluation propose des composantes invariables et universelles qui doivent se trouver à la base de toute démarche évaluative. En somme, la logique de l'évaluation, en tant que raisonnement fondamental, se veut une structure qui transcende les considérations méthodologiques liées au choix du modèle.

En conclusion, la logique de l'évaluation se veut une supra-structure qui sert de pilier aux conclusions évaluatives en structurant la démarche évaluative en une séquence logique et systématique. Ainsi, les modèles d'évaluation ne peuvent être confondus avec la logique d'évaluation. Ils représentent un paramètre périphérique à la logique qui vient influencer certains choix méthodologiques, mais qui n'affecte en rien la structure fondamentale nécessaire pour fonder un jugement et des conclusions évaluatives valides.

5. LES OBJECTIFS DE LA RECHERCHE

L'élaboration du contexte a permis de cerner l'ensemble des paramètres de la logique de l'évaluation et de les définir, ce qui s'avère essentiel pour répondre à la question de recherche sur la mise en application des paramètres de la logique de l'évaluation par les praticiens.

Plus précisément, cette recherche comporte les objectifs suivants :

1. Établir dans quelle mesure les praticiens de l'évaluation ont recours aux paramètres de la logique de l'évaluation pour structurer leur démarche évaluative;
2. Établir l'influence éventuelle de certaines variables, telles que la présence ou l'absence d'un cadre théorique de référence et l'utilisation de modèles d'évaluation, sur l'utilisation des paramètres et, possiblement, sur le niveau d'utilisation des paramètres;
3. Déterminer s'il y a une évolution dans les pratiques;
4. Déterminer dans quelle mesure les évaluateurs se réfèrent aux informations recueillies pour fonder leur jugement.

CHAPITRE III

MÉTHODOLOGIE

Tel que précisé antérieurement, la présente démarche a pour objet d'étude les pratiques évaluatives en ce qui a trait à l'utilisation des paramètres de la logique de l'évaluation. Ce chapitre abordera le type de recherche requis, les choix méthodologiques en termes d'échantillon, de collecte et d'analyse de données, et les limites de la recherche.

1. LE TYPE DE RECHERCHE

Cette recherche se fonde sur une démarche descriptive qui, comme le mentionne Van der Maren (1996), consiste à identifier les éléments et les relations statiques existant entre eux. Pour ce faire, une enquête sur les pratiques sera réalisée à partir de publications scientifiques. Ainsi, en raison du nombre important de cas analysés, la méthodologie mixte est la plus appropriée pour répondre aux objectifs de cette recherche.

2. LES CHOIX MÉTHODOLOGIQUES

La méthodologie de cette recherche s'inspire d'une étude réalisée par Lipsey (1988), qui constitue à ce jour un classique dans le domaine. L'auteur a réalisé une méta-analyse afin de générer, d'une part, une théorie concernant la contribution des schémas expérimentaux et, d'autre part, la validité des recherches et évaluations s'inscrivant dans le paradigme de la recherche quantitative-comparative (schémas

expérimentaux et quasi-expérimentaux). Cette étude s'avère un modèle quant à sa rigueur sur le plan du choix de l'échantillon et sur celui de la méthode de collecte et d'analyse des données. Considérant cependant que la présente recherche annonce des objectifs d'une autre teneur, certains choix appropriés ont été effectués, comme celui d'avoir recours à des données secondaires.

2.1 Le choix de l'échantillon

Considérant que la présente étude constitue la première phase d'un programme de recherche, il a été décidé qu'elle analyserait, dans un premier temps, des articles publiés dans des revues scientifiques en évaluation de programmes parce qu'ils constituent un critère de qualité en soi. Nous avons choisi de sélectionner trois revues sur une période de trois ans, soit un total de neuf années de revues. Les trois revues sélectionnées sont : la *Revue canadienne en évaluation de programme*, rattachée à la Société canadienne en évaluation; *The American Journal of Evaluation*, une publication rattachée à l'*American Evaluation Association*; et *Evaluation and Program Planning*, qui recueille une cote d'appréciation élevée auprès d'évaluateurs chevronnés.

Tout comme pour l'étude de Lipsey (1988), notre recherche étudie trois années de publication pour chacune des revues, mais des années qui ne sont pas consécutives, afin de tenter de dégager une tendance au sein d'un même cycle. En effet, considérant que les remises en question de la pratique ont souvent un cycle de dix ans, nous avons décidé d'étudier les articles publiés il y a dix ans, cinq ans et au cours de la dernière année, soit des textes de 1994, de 1999 et de 2004. Les articles retenus portent exclusivement sur des démarches évaluatives, ce qui exclut par le fait même les textes théoriques, d'opinions ou de vulgarisation, de même que les démarches de *monitoring* (indicateurs de performance), de « *benchmarking* » et d'évaluation basée sur le modèle logique.

2.2 La collecte des données

L'analyse documentaire des articles a été effectuée au moyen d'une grille élaborée à partir de termes de référence développés dans le cadre théorique, à savoir la problématique et la question d'évaluation, l'objectif et le type d'évaluation, l'objet d'évaluation, les critères et les variables observables, les standards et le jugement. Cette grille, présentée à l'appendice A, comprend trois types d'information :

1. Des éléments d'identification de l'article, tels que l'année et la revue;
2. Une définition de chacun des paramètres;
3. Une échelle de jugement à trois niveaux.

La grille reprend les paramètres présentés à la figure 2.3 et les définit de façon opérationnelle, lui assurant ainsi une validité de contenu. La grille a été élaborée par Madame Marthe Hurteau et validée sur le plan du contenu par Messieurs Jean Bélanger et Sylvain Houle, professeurs à l'UQAM et tous deux ayant une expertise reconnue dans le domaine de l'évaluation de programmes.

Une fois cette première série d'opérations réalisée, la grille a été préexpérimentée par l'auteur réalisant le présent projet et une de ses collègues afin de dégager un coefficient d'agrément interjuge. Les écrits reconnaissent qu'un coefficient d'agrément minimal de 0.80 doit être atteint afin de garantir la stabilité des jugements qui émaneront de la démarche. Cette condition a été assurée avec un coefficient de 0,85¹ atteint dans le cadre de la préexpérimentation.

¹ Le calcul du coefficient d'agrément interjuge est présenté à l'appendice B.

La collecte de données a donc commencé par l'application systématique de la grille à chacun des articles de l'échantillon. Soulignons que les premiers articles ont été révisés à la fin du processus de collecte de données afin de garantir une uniformité de traitement. Toutefois, les jugements émis sont demeurés stables.

Les données recueillies ont été compilées au moyen du logiciel informatique Excel, qui s'avère une excellente plate-forme pour exporter des données de recherche dans SPSS, un logiciel reconnu de traitement de données numériques qui a été retenu pour analyser les résultats de cette recherche.

3. L'ANALYSE ET L'INTERPRÉTATION DES DONNÉES

Cette étape de l'analyse des données est cruciale puisqu'elle vise à produire des résultats susceptibles de répondre aux objectifs de la recherche mentionnés dans le cadre théorique, à savoir :

- Établir le niveau d'utilisation des paramètres de la logique de l'évaluation par les praticiens de l'évaluation;
- Établir le niveau d'influence potentiel de certaines variables, telles que la présence ou l'absence d'un cadre théorique de référence et l'utilisation de modèles d'évaluation, sur l'utilisation des paramètres;
- Déterminer s'il y a une évolution en ce qui a trait au niveau d'utilisation des paramètres;
- Déterminer dans quelle mesure les évaluateurs se réfèrent aux informations recueillies pour fonder leur jugement.

Une analyse descriptive des données a été réalisée afin de dégager les différentes tendances observables au sein de la distribution des variables de cette recherche. Cette analyse sera développée dans le chapitre sur les résultats.

En ce qui a trait au premier objectif, l'analyse des résultats a permis de faire ressortir dans quelle proportion les évaluateurs utilisent les paramètres de la logique de l'évaluation pour structurer leur démarche. Le deuxième objectif, en plus de la compilation et de l'analyse descriptive, a été étudié en termes d'impact des autres variables sur le niveau d'utilisation des paramètres de la logique de l'évaluation par le calcul de coefficient khi carré. Pour le troisième objectif, la démarche utilisée pour l'objectif deux a été reprise, mais en privilégiant l'année de publication comme variable ayant un impact sur le recours aux paramètres de la logique de l'évaluation. Ce traitement a pour but de vérifier si le niveau d'utilisation des paramètres varie dans le temps à l'intérieur d'un même cycle de 10 ans de remise en question. Le dernier objectif a nécessité le recours à un traitement de nature qualitative, cette fois. En effet, une analyse de contenu a été réalisée pour établir si les évaluateurs fondent réellement leur jugement sur les informations recueillies et ainsi déterminer si le recours aux paramètres de la logique de l'évaluation permet d'obtenir des conclusions répondant à la question et aux objectifs d'évaluation.

En conclusion, l'ensemble de ces traitements a permis de produire un portrait descriptif de l'utilisation des paramètres de la logique de l'évaluation et a contribué à établir, dans la mesure du possible, l'impact de ces paramètres sur la qualité du jugement émis par les praticiens et, par le fait même, sur la validité des conclusions des évaluations.

4. LES LIMITES DE LA RECHERCHE

Toute recherche implique des contraintes et des choix méthodologiques qui restreignent la portée d'extrapolation ou de généralisation des résultats. La présente étude a donc aussi ses limites, qui sont présentées dans cette section.

La première limite réside dans le fait que seulement trois revues d'évaluation ont été retenues et qu'elles sont toutes nord-américaines. De plus, elles ne portent que sur les dix dernières années. Finalement, l'auteure du présent mémoire ne possède qu'une connaissance et une expérience limitées du champ théorique et de la pratique en évaluation de programme.

Bref, cette étude a produit des résultats rigoureux, mais dont la portée de généralisation ne peut aller au-delà des conditions de réalisation de la recherche.

CHAPITRE IV

RÉSULTATS

Dans un premier temps, il convient de détailler la nature de l'information issue de la présente démarche afin de la contextualiser. Ainsi, l'analyse documentaire porte sur 69 articles provenant des trois publications mentionnées dans la méthodologie, à savoir la *Revue canadienne en évaluation de programme*, l'*American Journal of Evaluation* et la revue *Evaluation and Program Planning*, et qui sont parus pendant les trois années déterminées, soit 2004, 1999 et 1994.

Le tableau 4.1 présente la répartition des articles retenus selon les revues et les années par rapport au nombre total de textes.

Tableau 4.1. Recension des articles analysés

		1994	1999	2004	Total
<i>Revue canadienne en évaluation de programme</i>	Retenus	2	4	4	10
	Total	15	16	14	45
<i>American Journal of Evaluation</i>	Retenus	2	3	7	12
	Total	15	40	33	88
<i>Evaluation and Program Planning</i>	Retenus	12	16	19	47
	Total	47	39	42	128
Total	Retenus	16	23	30	69
	Total	77	95	89	261

Seulement 25% des articles publiés au cours de 1994, 1999 et 2004 dans les revues ciblées ont été retenus pour les fins de notre analyse, les autres textes ne répondant pas aux critères énoncés préalablement. En effet, plusieurs des articles étaient plutôt des éditoriaux, des critiques de livre ou des textes d'opinion sur un sujet de l'actualité. Il est intéressant de constater que 68% des articles retenus proviennent de la revue *Evaluation and Program Planning*, qui semble publier le plus grand nombre de démarches d'évaluation. De plus en plus de publications, dont la *Revue canadienne en évaluation de programme*, exigent des apports théoriques ou conceptuels dans les articles qu'elles font paraître. Cette politique limite le nombre d'articles répondant à nos critères de sélection de l'échantillon. Enfin, on observe que 44% des articles retenus ont été publiés en 2004. Notons que l'*American Journal of Evaluation* a produit quatre numéros au lieu de trois, cette année-là.

Étant donné que les résultats de la recherche seront présentés en fonction des objectifs énoncés, il convient de rappeler en quoi consistent ceux-ci :

- Établir le niveau d'utilisation des paramètres de la logique de l'évaluation par les praticiens pour structurer leur démarche évaluative;
- Étudier l'influence de certaines variables, telles que la présence (ou l'absence) d'un cadre théorique de référence et l'utilisation de modèles d'évaluation, sur l'utilisation des paramètres et, possiblement, sur le niveau d'utilisation des paramètres;
- Déterminer s'il y a une évolution dans les pratiques;
- Établir dans quelle mesure les évaluateurs se réfèrent aux informations recueillies pour fonder leur jugement.

1. OBJECTIF 1 : LE RECOURS AUX PARAMÈTRES DE LA LOGIQUE DE L'ÉVALUATION

Une analyse descriptive des paramètres de la logique de l'évaluation, réalisée à partir des résultats consignés dans les grilles d'analyse, ainsi qu'une analyse des cas ont permis de générer les résultats reliés au premier objectif. Les différents tableaux inclus dans cette section dressent un portrait de la situation.

1.1 Les résultats descriptifs

Dans quelle mesure les praticiens ont-ils recours aux paramètres de la logique d'évaluation? Le tableau 4.2 présente une synthèse de l'information recueillie en établissant la fréquence pour chacun des paramètres de la logique de l'évaluation.

Tableau 4.2. Distribution des fréquences pour chaque paramètre

Paramètres	Absent	Non	En partie		Oui	Total
			Faible	Fort		
Description du programme			2 (3%)	20 (29%)	47 (68%)	69 (100%)
Problématique			5 (7%)	50 (73%)	14 (20%)	69 (100%)
Question présente		13 (19%)		44 (64%)	12 (17%)	69 (100%)
Question en lien avec la problématique	13 (19%)			5 (7%)	51 (74%)	69 (100%)
Question formulée clairement	13 (19%)			44 (64%)	12 (17%)	69 (100%)
Objectifs présents					69 (100%)	69 (100%)
Objectifs en lien avec la question				4 (6%)	65 (94%)	69 (100%)
Objectifs formulés clairement		9 (13%)			60 (87%)	69 (100%)
Objet présent					69 (100%)	69 (100%)
Objet en lien avec l'objectif					69 (100%)	69 (100%)
Critères présents					69 (100%)	69 (100%)
Choix des critères justifié			3 (4%)	59 (86%)	7 (10%)	69 (100%)
Critères formulés clairement		12 (17%)			57 (83%)	69 (100%)
Variables observables en lien avec les critères		3 (4%)	6 (9%)	35 (51%)	25 (36%)	69 (100%)
Standards présents		27 (39%)			42 (61%) Comp. Externe 32 10 (46%) (15%)	69 (100%)
Jugement basé sur des critères					69 (100%)	69 (100%)
Jugement basé sur des standards			27 (39%)		42 (61%)	69 (100%)
Jugement couvre les objectifs d'évaluation					69 (100%)	69 (100%)

Il ressort de ce tableau que la plupart des paramètres obtiennent une fréquence élevée en ce qui concerne leur présence dans les articles analysés. Les résultats pour chacun des paramètres sont étudiés plus en détail dans les prochaines sections.

1.1.1 La description du programme

L'auteur décrit brièvement le programme de façon à ce que le lecteur puisse en comprendre l'essence. (Libellé de la grille d'analyse.)

La description du programme est essentielle pour contextualiser les paramètres de la logique de l'évaluation. En ce sens, la description du programme est la mise en contexte de l'évaluation permettant de juger de la pertinence des choix stratégiques et méthodologiques des évaluateurs pour chaque évaluation analysée.

Concrètement, dans la grande majorité des cas, la description est suffisante pour comprendre en quoi consiste le programme sur lequel porte l'évaluation. En effet, 68% des articles présentent une description claire et complète, et 29% présentent une description suffisante mais succincte de leur programme. Seuls 3% des articles analysés comportent une description du programme jugée à peine suffisante pour en saisir l'essence.

De façon générale, la description est suffisante pour saisir ce en quoi consiste le programme. Les résultats pour ce paramètre ne font pas état d'une situation problématique susceptible d'affecter la qualité des autres paramètres.

1.1.2 La problématique

L'auteur présente les motifs ou raisons qui justifient la réalisation de l'évaluation. (Libellé de la grille d'analyse.)

La problématique cerne la raison ou le problème ayant motivé la réalisation de l'évaluation. Ce paramètre est des plus importants puisque l'ensemble des autres paramètres en découle. En ce sens, le jugement posé lors de l'évaluation tente de répondre à cette problématique. Il importe de bien la définir et d'y référer tout au long de la démarche afin de limiter les dérives et de s'assurer que les conclusions soient pertinentes.

Dans les articles retenus, la problématique est généralement bien présentée (93%). Toutefois, seulement 20% des problématiques font état de la justification proprement dite de l'évaluation du programme. Par ailleurs, 7% n'ont pas su démontrer le problème ou les raisons ayant motivé la réalisation de l'évaluation.

Enfin, la question et/ou les objectifs d'évaluation devraient être directement en lien avec la problématique. Les résultats démontrent qu'une problématique mal définie affecte la qualité des paramètres qui en découlent. En ce sens, lorsque la question est absente, c'est souvent parce que la problématique est imprécise ou absente.

1.1.3 La ou les questions

L'auteur formule la ou les questions auxquelles l'évaluation doit apporter des éléments de réponse. La question doit être en lien avec la problématique. (Libellé de la grille d'analyse.)

La question d'évaluation opérationnalise la problématique et annonce les objectifs. Ce paramètre prend tout son sens lorsqu'il s'agit de vérifier la cohérence de la démarche. En ce sens, chaque paramètre devrait contribuer à répondre à la question.

La question d'évaluation est présente dans 17% des évaluations analysées. La question présentée implicitement sous forme d'affirmation se retrouve dans 64% des textes, et elle est proprement absente dans 19% des cas. De façon générale, la question est très en lien avec la problématique (74%), ou partiellement en lien (7%). Dans le reste des cas (19%), soit ceux où la question est absente, il n'est pas pertinent d'établir des conclusions sur le lien entre la question et la problématique. Par ailleurs, la question est clairement formulée sous forme d'interrogation dans seulement 17% des évaluations analysées. Elle est présente sous forme d'affirmation dans 64% des cas. Évidemment, lorsque la question était absente, c'est-à-dire dans 19% des articles étudiés, il ne s'est pas avéré pertinent de parler de sa clarté.

En résumé, lorsque la question est présente, elle est généralement en lien avec la problématique et elle est clairement formulée. Lorsqu'elle est absente, il n'y a pas lieu de statuer sur sa clarté ni sur son lien avec la problématique.

1.1.4 L'objectif d'évaluation et/ou le type d'évaluation

L'auteur énonce les objectifs de l'évaluation selon la typologie de Rossi, Lipsey et Freeman (2004). L'objectif doit être en lien avec la question ou la problématique et être formulé clairement. (Libellé de la grille d'analyse.)

L'objectif d'évaluation est un paramètre central dans la logique de l'évaluation, affectant directement la détermination de l'objet, des critères et des standards. Par conséquent, il importe de s'assurer qu'il est en lien avec la question d'évaluation et la problématique et que sa formulation est suffisamment claire pour définir les paramètres qui en découlent.

Dans les évaluations analysées, l'objectif est généralement très bien circonscrit. L'objectif est présent dans toutes les évaluations analysées (100%) et il est fortement en lien avec la problématique dans 94% des textes. Il est formulé clairement dans 87% des cas. L'évaluation la plus fréquemment réalisée est l'évaluation d'impact (64%), mais on retrouve aussi régulièrement des évaluations d'implantation (13%) et de pertinence (12%). Occasionnellement, les articles font état d'évaluations de coût-bénéfice (7%) et de conceptualisation (4%). La typologie de Rossi, Freeman et Lipsey est utilisée explicitement dans la plupart des articles analysés.

En somme, les objectifs sont bien définis dans les évaluations étudiées. Dans les évaluations où les objectifs ne sont pas formulés clairement, il ressort que les critères sont souvent aussi mal définis et que les standards sont absents.

1.1.5 L'objet d'évaluation

L'auteur identifie la dimension du programme sur laquelle va porter le jugement. La dimension doit être en lien avec l'objectif. (Libellé de la grille d'analyse.)

L'objet d'évaluation cible avec précision le ou les aspects des programmes qui sont à l'étude. Avant de se lancer dans la collecte de données, il importe de s'assurer que ce paramètre est cohérent avec les objectifs et que son énoncé circonscrit avec précision l'étendue de l'évaluation.

Concrètement, l'objet est toujours présent et très lié à l'objectif de l'évaluation (100%). D'ailleurs, dans les textes analysés, l'objet est généralement présenté avec l'objectif au sein d'un même paragraphe.

1.1.6 Les critères

L'auteur identifie les points de référence sur lesquels il a basé son jugement. Le choix des critères est justifié. Les critères sont formulés clairement. (Libellé de la grille d'analyse.)

Les critères servent à traduire les objectifs en éléments précis sur lesquels portera la collecte de données. Afin de maintenir un niveau élevé de cohérence, il est important que les critères soient congruents aux objectifs.

Dans les textes analysés, les critères sont toujours présents (100%), mais seulement 10% des articles justifient le choix de leurs critères. Dans 86% des cas, la justification est implicite et liée au type d'évaluation effectuée. Les critères sont formulés clairement dans 83% des cas alors que, dans les 17% restants, ils sont trop sous-entendus pour être identifiés avec certitude.

Notons que, lorsque les critères ne sont pas mentionnés clairement, les variables observables sont aussi formulées moins clairement et qu'il est difficile de les rattacher aux critères.

1.1.7 Les variables observables

L'auteur identifie clairement les différentes variables observables qui serviront à mesurer les critères. (Libellé de la grille d'analyse.)

Les variables observables opérationnalisent les critères et constituent les éléments concrets sur lesquels se fonde la mesure. Ce paramètre se doit d'être très congruent aux critères qu'il entend mesurer.

Les variables observables sont clairement associées aux critères dans 36% des évaluations étudiées et le sont implicitement dans 51% des cas. Dans 9% des articles, les variables sont à peine associées aux critères alors que, dans 4% des cas, elles ne le sont pas du tout.

Tel que mentionné plus haut, lorsque les critères ne sont pas clairement formulés, les variables observables ne le sont pas non plus. Dans ce cas, il convient de s'interroger sur la validité des résultats qui en découlent.

1.1.8 Les standards

L'auteur identifie des standards pour chaque critère ou variable observable. Ils doivent être en lien avec les variables observables ou les critères. (Libellé de la grille d'analyse.)

Les standards sont un paramètre important dans la logique de l'évaluation : ils servent à établir dans quelle mesure l'objet de l'évaluation rejoint les attentes ou objectifs du programme. Le jugement devrait se fonder sur les écarts observés entre la mesure des variables observables et les standards, établis idéalement *a priori*.

Les évaluateurs ont eu recours à des standards dans 61% des cas. Cependant, il est à noter que seulement 15% des études ont recours à des standards externes ou absolus comparativement à 46% qui utilisent des standards comparatifs fondés sur des groupes témoins ou des données pré-test.

1.1.9 Le jugement

L'auteur démontre que son jugement est établi à partir des écarts entre les observations et les critères, standards. (Libellé de la grille d'analyse.)

L'auteur démontre que l'évaluation a répondu à la question d'évaluation et couvert l'objectif de l'évaluation. (Libellé de la grille d'analyse.)

Finalité de l'évaluation, le jugement sert à statuer sur la valeur d'un programme ou d'un ou plusieurs de ses aspects. La crédibilité des conclusions émises au terme d'une évaluation dépend de la cohérence de l'ensemble des paramètres de la logique de l'évaluation. Il importe donc que le jugement réponde aux objectifs et à la question d'évaluation.

Dans l'ensemble des cas analysés, le jugement se fonde sur les critères établis (100%). Dans 61% des cas, il s'appuie sur des standards alors que, dans 39% des cas, il se fonde sur des faits non appuyés par des standards. Il appert que ces deux paramètres, soit le jugement et les standards, sont très liés. En effet, si les jugements s'appuient sur le recours à des standards dans 61% des cas, il semblerait que les jugements, dans les autres 39%, ont de la difficulté à dépasser le simple niveau des faits.

Les évaluations analysées ont toutes répondu à la question d'évaluation ou à la problématique de l'évaluation (100%). On comprend ainsi que le jugement couvre l'objectif de l'évaluation dans l'ensemble des cas étudiés.

En conclusion, les articles analysés font état des paramètres de la logique de l'évaluation dans la quasi-totalité des cas. La logique de l'évaluation semble donc être sous-jacente à la plupart des évaluations étudiées. La section suivante présente les résultats article par article afin de dégager des tendances globales.

1.2 L'analyse des cas

Il est possible de dégager certaines tendances générales au sein des résultats, bien que les tests statistiques d'association (χ^2) ne soient pas significatifs. Le tableau suivant permet d'observer ces tendances en présentant les résultats selon la présence ou l'absence du paramètre dans l'évaluation, sans égard aux nuances présentées dans la grille (« En partie fort » et « En partie faible »). Les fréquences obtenues dans la colonne « En partie fort » ont été regroupées avec les réponses appartenant au « Oui », tandis que les résultats de la colonne « En partie faible » ont été regroupés avec les réponses appartenant au « Non ».

Il ressort de ces tableaux que les paramètres identifiés dans la logique de l'évaluation sont présents dans la quasi-totalité des évaluations étudiées. Comme les paramètres de la logique de l'évaluation sont étroitement liés, l'imprécision ou l'absence de l'un d'entre eux risque d'avoir un impact sur la présence et la qualité des autres. C'est pourquoi certains paramètres demandent quelques nuances.

La clarté de la question d'évaluation est l'une des variables qui obtient le taux de présence le plus faible, soit 17%. Ce résultat couvre deux réalités. D'une part, dans 19% des cas, la question est absente; il n'est donc pas pertinent de statuer sur la clarté de la formulation de la question. D'autre part, dans les 64% restants, la question est présente, mais elle n'est pas formulée clairement, prenant la forme d'une affirmation et non d'une interrogation. En somme, bien que la clarté semble problématique de prime abord, ce paramètre ne pose pas réellement problème en ce qui a trait à la logique d'évaluation présentée dans cette recherche.

On observe que lorsque les objectifs sont formulés de façon implicite, les critères le sont souvent aussi. Quant aux variables observables, leur lien avec les critères est dans certains cas trop imprécis pour être établi avec certitude; cette situation survient surtout lorsque les critères ne sont pas formulés clairement. Il devient alors difficile de rattacher les variables à des critères qui sont eux-mêmes difficilement identifiables. Par contre, il importe de noter que cette tendance n'engendre pas une situation problématique par rapport à ces paramètres.

Par ailleurs, une étude plus approfondie du tableau 4.3 permet d'identifier des profils d'évaluation, c'est-à-dire que certaines évaluations se conforment davantage à la logique de l'évaluation que d'autres. Il n'est pas possible de déterminer si l'absence du paramètre est due au fait que l'évaluateur ne l'a pas défini dans son évaluation ou si cette absence est due à une omission dans la rédaction. Pour les fins de cette recherche, nous prenons pour acquis que, lorsqu'un paramètre est défini dans l'évaluation, le praticien a pris soin d'en faire mention dans son article. Le tableau suivant résume les résultats du tableau 4.3 en présentant le nombre de variables absentes par rapport au 17 variables identifiées dans la grille.

Tableau 4.4. Profil d'évaluations

Cases en gris : L'évaluation répond parfaitement ou presque parfaitement à la logique de l'évaluation (moins de 4 absences).

Cases en blanc : L'évaluation présente quelques lacunes par rapport à la logique de l'évaluation (4 à 8 absences).

ID Article	Nombre d'absences	ID Article	Nombre d'absences	ID Article	Nombre d'absences
EPP_04_1	1	AJE_04_1	5	CJE_99_1	5
EPP_04_2	1	AJE_04_2	3	CJE_99_2	5
EPP_04_3	1	AJE_04_2	7	CJE_99_3	3
EPP_04_4	5	AJE_04_2	1	CJE_99_4	2
EPP_04_5	3	AJE_04_2	1	AJE_99_1	0
EPP_04_6	1	AJE_04_2	4	AJE_99_2	1
EPP_04_7	1	AJE_04_2	1	AJE_99_3	3
EPP_04_8	1	EPP_99_1	3	EPP_94_1	6
EPP_04_9	3	EPP_99_2	1	EPP_94_2	1
EPP_04_10	3	EPP_99_3	2	EPP_94_3	0
EPP_04_11	3	EPP_99_4	2	EPP_94_4	1
EPP_04_12	2	EPP_99_5	3	EPP_94_5	6
EPP_04_13	4	EPP_99_6	0	EPP_94_6	2
EPP_04_14	1	EPP_99_7	3	EPP_94_7	5
EPP_04_15	1	EPP_99_8	1	EPP_94_8	1
EPP_04_16	1	EPP_99_9	0	EPP_94_9	2
EPP_04_17	7	EPP_99_10	2	EPP_94_10	6
EPP_04_18	1	EPP_99_11	2	EPP_94_11	8
EPP_04_19	3	EPP_99_12	0	EPP_94_12	2
CJE_04_1	5	EPP_99_13	0	CJE_94_1	5
CJE_04_2	3	EPP_99_14	4	CJE_94_2	1
CJE_04_3	1	EPP_99_15	0	AJE_94_1	3
CJE_04_4	3	EPP_99_16	1	AJE_94_2	6

Il semble se dégager de l'étude du tableau que les évaluations ayant moins de 4 variables absentes, c'est-à-dire celles qui se conforment à la logique de l'évaluation, représentent 75% des cas analysés. Pour ce qui est des autres 25% des évaluations, elles offrent un profil s'éloignant d'une façon plus ou moins importante de la logique de l'évaluation de cette recherche selon le nombre de variables absentes (entre 4 et 8). Malgré tout, nous remarquons que ces évaluations présentent une démarche qui suit une certaine structure à laquelle il est possible de se référer pour valider les conclusions évaluatives.

D'un autre côté, comme ce tableau tente d'établir la qualité de la démarche évaluative et que cette dernière semble être très liée au recours à des standards, il convient de préciser que le type d'évaluation semble avoir un impact sur le recours aux standards, soulevant même la question de la faisabilité et de la pertinence de recourir à ceux-ci dans certains cas. Le tableau suivant illustre cette tendance.

Tableau 4.5. Tableau croisé entre le type d'évaluation et les standards*Cases en gris : Tendances observées.*

		Standards			Total
		Oui : Standards externes	Oui : Standards comparatifs	Non : Standard absent	
Type d'évaluation	Conceptualisation	0	1	2	3
	Coût-bénéfice	3	1	1	5
	Impact	7	29	8	44
	Implantation	0	1	8	9
	Pertinence	0	0	8	8
Total		10	32	27	69

On observe ici que les évaluations de pertinence sont celles qui recourent le moins aux standards (0%), alors que les évaluations d'impact et de coût-bénéfice y ont recours dans la très grande majorité des cas (82% et 80%). Cette constatation n'est pas vraiment surprenante si l'on considère la nature des différents types d'évaluation. En effet, comme l'évaluation de pertinence implique de dégager les besoins des utilisateurs du programme et de vérifier dans quelle mesure les activités y répondent, la démarche utilisée pour ce type d'évaluation a souvent un caractère exploratoire, ce qui rend l'établissement de standards *a priori* moins adéquat, voire difficilement faisable. Et c'est sans compter que la littérature ne fournit pas toujours l'information nécessaire pour constituer un cadre théorique de référence bien défini pour fonder les critères et les standards. Par contre, lorsqu'il s'agit d'une évaluation d'impact, qui

s'inscrit dans une démarche plus vérificatoire, il est important de déterminer les standards dès le début de la démarche afin de s'assurer que les résultats rencontrent réellement les objectifs du programme. D'ailleurs, les standards devraient idéalement être contenus dans le libellé des objectifs du programme. Par choix, ou parce que la littérature ne permet pas d'établir un cadre théorique de référence solide pour fonder les critères et les standards, il arrive que les évaluateurs aient recours à des comparaisons entre un groupe expérimental et un groupe témoin plutôt qu'à des standards établis *a priori*. Enfin, ces résultats quant au lien entre le type d'évaluation et la présence des standards nous permettent de nuancer les résultats plutôt faibles concernant le taux d'utilisation des standards, qui étaient de 61% : parmi les 39% d'évaluations n'ayant pas fait état de standards, 12% sont des évaluations de pertinence où il n'est pas toujours adéquat ou faisable d'en établir.

En conclusion, l'étude des résultats concernant la présence des paramètres permet de dégager une présence importante des paramètres de la logique de l'évaluation tels que présentés dans ce mémoire. Ainsi, il est possible de conclure à une utilisation importante de la logique de l'évaluation au sein des démarches évaluatives étudiées, validant, de ce fait, les paramètres proposés à la figure 2.3 (p. 17), du moins en ce qui concerne les évaluations ayant fait l'objet d'une publication.

2. OBJECTIF 2 : L'INFLUENCE D'AUTRES VARIABLES

Cet objectif de recherche vise à valider certaines conclusions émises par d'autres auteurs concernant des éléments périphériques à la logique de l'évaluation, soit la présence ou l'absence d'un cadre théorique de référence en soutien à l'évaluation et le recours ou l'absence de recours à un modèle d'évaluation en soutien à la démarche évaluative, étudié par Christie (2003). Tel que mentionné dans le cadre théorique, Lipsey (1988) a démontré dans son étude que les praticiens font peu appel à des

cadres théoriques de référence pour situer le contexte théorique de leur évaluation, bien que le recours à un contexte théorique confère plus de validité aux critères. Cette recherche tente de déterminer si le recours à un cadre théorique de référence pour établir les critères de l'évaluation a un impact sur la présence de la logique de l'évaluation comme structure de l'évaluation dans les articles étudiés. Par ailleurs, Christie (2003) a établi que les praticiens de l'évaluation recourent peu aux modèles de l'évaluation développés précisément pour les aider à structurer leur démarche évaluative. Il est intéressant de vérifier dans quelle mesure le recours à des modèles d'évaluation peut influencer l'utilisation de la logique de l'évaluation comme structure de l'évaluation et être compatible avec cette utilisation. Le tableau suivant présente les résultats obtenus pour ces variables en termes de fréquences.

Tableau 4.6. Résultats pour les variables périphériques

Variables	Non	Oui	Total
Cadre théorique de référence	3	66	69
Modèle d'évaluation	65	4	69

2.1 La présence d'un cadre théorique de référence

L'auteur présente la théorie sous-jacente à la démarche (présence d'un cadre théorique de référence). (Libellé de la grille d'analyse.)

Dans la quasi-totalité des évaluations, soit dans 66 évaluations ou 96% des cas étudiés, il est fait mention d'un cadre théorique en support à l'évaluation. Cette forte majorité fait en sorte qu'il n'est pas pertinent de pousser plus loin l'analyse de cet élément.

2.2 Le recours à un modèle d'évaluation

L'auteur mentionne le recours à un modèle d'évaluation reconnu dans les écrits et il en justifie l'utilisation. (Libellé de la grille d'analyse.)

Seulement 4 évaluations sur les 69 analysées affirment avoir eu recours à un modèle d'évaluation pour structurer leur évaluation. De ces 4 cas, une évaluation mentionne un modèle axé sur la méthode, une un modèle axé sur le jugement et deux un modèle axé sur la prise de décision. Ce faible résultat va dans le même sens que l'étude de Christie (2003), qui note que les évaluateurs ont peu recours à des modèles d'évaluation pour structurer leur démarche, et rend non pertinente une analyse plus poussée.

En conclusion, il est difficile de dégager des tendances ou des liens entre ces variables périphériques et les paramètres de la logique de l'évaluation en raison de la présence très fréquente du contexte théorique et de l'absence marquée des modèles d'évaluation.

3. OBJECTIF 3 : L'ÉVOLUTION DANS LES PRATIQUES

Les résultats présentés au tableau 4.3 (p. 47) permettent de dégager certains constats sur l'évolution de la pratique en ce qui concerne le recours à la logique de l'évaluation, bien que les tests statistiques d'association (χ^2) ne permettent pas de les confirmer. De cette analyse, il est possible de dégager les points suivants :

1. Les questions, en 1999, sont presque toujours présentes, soit dans 22 cas sur 23, alors que, pour les deux autres années, 6 évaluations par année ne présentent pas de question.
2. Il en va de même pour les objectifs : en 1999, ils sont formulés clairement dans tous les cas sauf un alors que, pour les deux autres années, 4 articles par année présentent des objectifs dont la formulation est imprécise.
3. En 2004, il y a moins de critères formulés indistinctement que pour les deux autres années, soit 10% (3/30) en 2004 pour 13% (3/23) en 1999 et 31% (5/16) en 1994.
4. Quant aux standards, leur présence dans les évaluations est plutôt constante au fil des ans, soit 56% en 1994, 65% en 1999 et 60% en 2004.

Ensuite, l'analyse des résultats présentés dans le tableau 4.5 (p. 53) permet de dégager une tendance similaire quant à l'amélioration des évaluations avec le temps. En effet, en 1994, 44% des évaluations, soit 7 évaluations sur 16, font état de 4 variables absentes ou plus, alors qu'il n'y en a plus que 13% en 1999 (3/23) et 23% en 2004 (7/30). Cette tendance illustre que la logique de l'évaluation est une structure de plus en plus employée dans les évaluations publiées dans les revues analysées, ou alors elle révèle que les praticiens utilisent davantage le modèle de présentation des données imposé par la rédaction de ces revues.

En conclusion, il ressort que les profils des évaluations sont plutôt similaires à travers les années analysées. Bien qu'il ne soit pas possible de conclure à une évolution des pratiques, la lecture des articles fait ressortir que les paramètres de la logique de l'évaluation ne semblent plus simplement présents, mais sont de plus en plus intégrés à la pratique évaluative au fil des ans.

4. OBJECTIF 4 : LE FONDEMENT DU JUGEMENT

Ce dernier objectif a un caractère plus qualitatif. Il s'agit d'établir dans quelle mesure le jugement se fonde effectivement sur les informations recueillies au cours de l'évaluation. Pour ce faire, il faut juger de la qualité des conclusions émises dans l'évaluation en regard des informations recueillies.

Le jugement est ce qui permet d'établir des conclusions quant à la problématique et aux objectifs de l'évaluation. C'est un acte professionnel que les évaluateurs accomplissent à chaque mandat. Comme Scriven (1995) le mentionne, la crédibilité du jugement est étroitement liée à la cohérence entre les différents éléments de la logique de l'évaluation.

De façon générale, les évaluations étudiées sont bien construites et présentent des conclusions qui semblent congruentes à la démarche évaluative et aux informations ainsi obtenues. Tel que mentionné précédemment, il est possible de dégager des profils d'évaluation qui semblent largement influencés par le type d'évaluation en cause et qui semblent conditionner l'importance de certains paramètres par rapport aux autres. Par exemple, les évaluations de pertinence accordent en général moins d'importance aux standards et fondent leur jugement surtout sur des faits, alors que les évaluations d'impact et de coût-bénéfice utilisent presque toujours des standards externes pour fonder leur jugement.

Enfin, le jugement se base majoritairement sur des critères et des standards. Dans les évaluations où les standards sont absents, le jugement est supposé s'appuyer sur des faits. Cette situation apparaît moins valable que la première. Par ailleurs, le jugement présenté dans les évaluations étudiées couvrait toujours le ou les objectifs d'évaluation en cause. Le recours à la logique de l'évaluation permet au moins de garantir cette cohérence dans la démarche.

5. LA SYNTHÈSE DES RÉSULTATS

Les résultats de cette recherche mettent en relief un certain nombre de constats. Tout d'abord, en ce qui concerne l'objectif 1, il ressort que les paramètres de la logique de l'évaluation se retrouvent dans la majorité des cas étudiés. Ensuite, pour l'objectif 2, il se dégage certaines tendances quant à l'impact de l'utilisation d'un cadre théorique de référence et d'un modèle d'évaluation. Le recours quasi général à un cadre théorique pourrait expliquer, partiellement du moins, la très grande utilisation des paramètres de la logique de l'évaluation en facilitant le choix des critères et des standards. Par contre, le faible taux d'utilisation de modèles d'évaluation par les praticiens semble démontrer que le recours aux modèles n'est pas nécessaire pour faire appel à la logique de l'évaluation comme soutien à la démarche évaluative. En ce qui a trait à l'objectif 3, la tendance observée suggère que l'utilisation de la logique de l'évaluation pour structurer la démarche évaluative croît avec le temps. Enfin, pour l'objectif 4, le jugement posé dans les évaluations analysées est généralement très cohérent avec les informations recueillies. La logique de l'évaluation soutient la plupart des évaluations étudiées, et les jugements découlant des démarches fondées sur cette structure sont congruents aux visées et objectifs des évaluations en cause.

CHAPITRE V

DISCUSSION

Les résultats de la recherche permettent de mettre en lumière certains constats quant à l'utilisation de la logique de l'évaluation par les praticiens.

- Les paramètres de la logique de l'évaluation sont présents dans la majorité des évaluations analysées.
 - L'utilisation de standards s'avère le paramètre qui affiche la fréquence d'apparition la plus faible.
 - Lorsque la question est absente, les standards le sont souvent aussi.
 - Le type d'évaluation semble avoir un impact sur la présence des standards.
- Certaines évaluations se conforment de plus près que d'autres à la logique de l'évaluation en omettant de 0 à 8 variables dans leur profil (tableau 4.5, p. 53).
- Les évaluations analysées font état d'un cadre théorique de référence et d'une absence de recours aux modèles d'évaluation dans la quasi-totalité des cas étudiés.
- Les profils des évaluations analysées s'améliorent au sein du cycle de 10 ans étudié.
- Le jugement se fonde sur les informations recueillies et répond à l'objectif d'évaluation dans tous les cas étudiés.

Au-delà de ces constats, les résultats ont aussi soulevé certaines questions quant aux limites de l'interprétation qu'il est possible d'en tirer. Les questions suivantes seront développées dans les prochains paragraphes :

1. Le recours à la logique de l'évaluation est-il réel ou s'agit-il d'une exigence éditoriale?
2. Les cadres théoriques annoncés dans les articles sont-ils réellement utilisés pour établir les critères et les standards?
3. Le faible recours à des standards remet-il en cause l'utilisation effective des cadres théoriques annoncés ?

En ce qui concerne le premier point, il ressort que la majorité des évaluations relatées dans les articles semble se conformer en tous points à la séquence proposée à la figure 2.3 (p. 17), et il semblerait que le phénomène s'amplifie au fil du temps. Ces deux constats nous amènent à soulever la question suivante : le recours grandissant à cette structure logique s'explique-t-il par son efficacité à produire des résultats crédibles et des conclusions valides? Il ne faudrait pas conclure trop promptement par l'affirmative. En effet, deux réserves peuvent être émises. D'une part, il existe une certaine mise en forme standard exigée dans la présentation des articles par la rédaction des revues scientifiques, surtout en ce qui concerne la revue *Evaluation and Program Planning*, qui suggère que le rapport de la démarche pourrait relever d'une exigence éditoriale. Il n'en demeure pas moins que les auteurs avaient le matériel requis pour s'y conformer, mais en aucun cas peut-on affirmer que cette « logique » leur a servi de cadre de référence pour élaborer et guider leur démarche. D'autre part, l'enquête n'ayant pas été effectuée auprès des auteurs, qui sont eux-mêmes des praticiens, il n'est pas possible de dissiper ce doute. Néanmoins, il demeure que cette logique est trop constante pour ne pas être considérée comme une pratique exemplaire, d'autant plus qu'elle est constatée dans des revues scientifiques.

En ce qui concerne le second point, à savoir le cadre théorique, il semblerait que les résultats obtenus sont passablement différents de ceux obtenus par Lipsey (1988) puisqu'on identifie un cadre théorique dans 96% des cas. Cette situation peut

s'expliquer par le fait que les récentes technologies (ex : informations disponibles via Internet) ont permis aux chercheurs de créer des réseaux et de s'appuyer sur des informations qui n'étaient pas disponibles auparavant. Néanmoins, rien ne permet de s'assurer que ces cadres théoriques ont réellement été utilisés pour établir les critères et les standards de l'évaluation. Ainsi, une série de références dans une publication scientifique ne garantit pas qu'elles ont été utilisées. Par contre, il s'agit d'un excellent point de départ en ce qui a trait à la crédibilité de l'évaluation.

En ce qui concerne le troisième point, il est effectivement étonnant de constater le faible recours à des standards. Les résultats élevés quant au recours à des cadres théoriques suggèrent que les praticiens ont les ressources pour déterminer des standards valides pour leurs évaluations. Or, 39% des évaluations étudiées n'ont pas établi de standards pour fonder leur jugement. Ce faible taux de présence des standards, dont seulement 15% sont des standards externes établis *a priori*, remet en cause le recours effectif à la littérature annoncée dans le cadre théorique de référence. Par contre, comme on ne peut établir si les cadres théoriques annoncés contenaient de l'information susceptible de guider l'établissement des standards, ce faible taux de recours aux standards ne constitue pas un élément suffisant pour conclure à la non utilisation des cadres théoriques de référence.

Finalement, la faible utilisation des modèles d'évaluation dans les évaluations analysées va dans le même sens que les résultats de Christie (2003). Il est possible que les auteurs des articles aient passé sous silence le modèle d'évaluation utilisé parce que les modèles s'intéressent davantage à la nature des relations avec le client et s'éloignent des aspects plus scientifiques de la démarche. Quoi qu'il en soit, Christie (2003) suggère que ce faible recours aux modèles d'évaluation s'explique par le fait que les évaluateurs s'avèrent souvent des « experts accidentels » qui exercent la fonction principale d'évaluateur de programme bien que leur formation relève d'une discipline connexe. Dans tous les cas, toutes ces remarques demeurent des hypothèses qu'il faudrait vérifier au moyen d'une enquête.

CONCLUSION

La présente étude visait à décrire dans quelle mesure les praticiens ont recours à la logique de l'évaluation pour structurer leur démarche évaluative. Il ressort de cette recherche que la logique de l'évaluation est présente dans la plupart des évaluations étudiées. On la retrouve à chaque étape de la démarche évaluative en soutien aux choix méthodologiques. Par ailleurs, le recours à la logique de l'évaluation semble assurer une certaine cohérence entre les paramètres, permettant ainsi de structurer la démarche. Néanmoins, il demeure que l'étude n'a porté que sur des évaluations ayant fait l'objet d'une publication dans l'une des revues scientifiques spécialisées dans le domaine ce qui, par le fait même, ne constitue pas un échantillon représentatif de la pratique.

Notre recherche constitue un premier pas dans l'étude de l'impact de la logique de l'évaluation sur la validité des démarches évaluatives en faisant ressortir que les productions des praticiens chevronnés semblent effectivement s'inscrire à l'intérieur de ces paramètres. Il convient, néanmoins, de pousser plus avant les recherches avant de considérer la logique de l'évaluation comme un pilier théorique de la pratique évaluative. Pour conclure, il serait intéressant de reprendre cette recherche en utilisant un échantillon d'évaluations n'ayant pas fait l'objet d'une publication afin de vérifier si elles ont recours à la logique de l'évaluation dans une proportions différentes que les évaluations publiées et étudiées dans le cadre de cette recherche. En effet, les nombreux rapports d'évaluation non publiés ne sont pas soumis au regard extérieur des confrères de l'évaluateur, qui doivent juger de la valeur et de l'intérêt de la démarche pour une publication scientifique. Les résultats pourraient ouvrir des pistes de réponse à la problématique soulevée dans cette recherche, soit l'incapacité de la pratique évaluative à produire des conclusions valides et à les mettre en oeuvre.

Par ailleurs, une recherche exploratoire sur les pratiques évaluatives et la satisfaction des clients par rapport aux conclusions évaluatives pourrait nous éclairer davantage sur les insatisfactions spécifiques des clients et les raisons de celles-ci. En effet, nombre d'hypothèses sont émises sur les causes des remises en question de l'évaluation. Une recherche auprès des consommateurs d'évaluation de programme (« *stakeholders* ») pourrait fournir des réponses précises à partir desquelles la pratique pourrait s'ajuster.

Malgré des résultats plutôt encourageants par rapport au recours, par les praticiens, à une théorie fondée sur la logique de l'évaluation pour structurer leur démarche évaluative, l'évaluation de programme et la validité des conclusions qu'elle produit continuent d'être remises en question. Il faudra chercher ailleurs les causes de la non application des conclusions évaluatives.

RÉFÉRENCES

- Alkin, M.C. 1972. « Evaluation Theory Development ». In *Evaluation Action Programs*, Weiss, C. (éd.), p. 105-117. Boston : Allyn & Bacon.
- Alkin, M. C. 1990. *Debate on Evaluation*, Newbury Park: Sage Publications.
- Alkin, M.C. et C.A. Christie. 2004. « An Evaluation Theory Tree ». In *Evaluation Roots : Tracing Theorists' Views and Influences*, sous la dir. de M.C. Alkin, p. 12-65. Thousand Oaks (Calif.) : Sage Publications.
- Boruch, R.F. et H. Gomez. 1977. « Sensitivity Bias and Theory in Impact Evaluation ». *Professional Psychology*, vol. 8, n° 4, p. 411-434.
- Campbell, D.T. 1957. « Factor Relevant to the Validity of Experiment in Social Settings ». *Psychological Bulletin*, vol. 54, p. 297-312.
- Chen, H.T. et P. Rossi. 1983. « Evaluating With Sense : The Theory-driven Approach ». *Evaluation Review*, vol. 7, p. 283-302.
- Chen, H-T. 2005. *Practical Evaluation Program*, Thousand Oaks: Sage Publications.
- Christie, C.A. 2003. « The Practice-Theory Relationship in Evaluation ». *New Direction for Evaluation*, n° 97, p. 1-93.
- Cook, T. et D. Campbell. 1976. « The Design and Conduct of Quasi-experiments and True Experiment in Field Setting ». In *Handbook of Industrial and Organizational Psychology*, sous la dir. de D. M. Dunnette, p. 223-326. Chigago : Rand McNally.
- Cousins, B.J. et L.M. Earl. 1992. « The Case for Participatory Evaluation ». *Educational Evaluation and Policy Analysis*, vol. 14, n° 4, p. 397-418.
- Cronbach, L.J. et R.E. Snow. 1977. *Aptitudes and Instructional Methods : A Handbook for Research on Interactions*. New York : Irvington.
- Dubois, N. et R. Marceau. 2005. « Un état des lieux théoriques de l'évaluation : une discipline à la remorque d'une révolution qui n'en finit pas ». *The Canadian Journal of Program Evaluation*, vol. 20, n° 1, p. 1-36.
- Eisner, E. 1976. « Educational Connoisseurship and Criticism : Their Form and Function in Educational Evaluation ». *Journal of Aesthetic Evaluation or Education*, vol. 10, p. 135-150.

- Fetterman, D.M. 1984. « Ethnography in Educational Research : The Dynamics of Diffusion ». In *Ethnography in Educational Evaluation*, Fetterman, D.M., p. 21-35. Beverly Hills (Calif.) : Sage Publications.
- Fournier, D.M. 1995. « Establishing Evaluative Conclusions : A Distinction Between General and Working Logic ». *New Direction for Evaluation*, n° 68, p. 15-32.
- Gaudreau, L. 2001. *Évaluer pour évoluer : les indicateurs et les critères*, Montréal : Logiques.
- Guba, E.G. 1972. « The Failure of Educational Evaluation ». In *Evaluating Action Programs : Readings in Social Action and Education*, sous la dir. de C. H. Weiss, p. 251-266. Boston : Allyn & Bacon.
- Guba, E.G. et Y.S. Lincoln. 1989. *Fourth Generation Evaluation*. Newbury Park (Calif.) : Sage Publications.
- House, E.R. 1980. *Evaluating With Validity*. Thousand Oaks (Calif.) : Sage Publications.
- Hurteau, M. 1991. *Strategic Choices in Programs Evaluation*, sous la dir. de A.E. Love, p. 106-125. Ottawa : Publications de la Société canadienne en évaluation de programme.
- _____. 2005. *Comprendre les pratiques en évaluation de programme afin de les améliorer : un retour aux fondements théoriques*. Projet de recherche subventionné dans le cadre du programme d'aide financière à la recherche et à la création : Université du Québec à Montréal.
- Hurteau, M. et S. Houle. (2005). « Évaluation et recherche évaluative : différences et similitudes ». *La revue de mesure et évaluation en éducation*
- King J.A. et E. Pechman. 1982. *Improving Evaluation Use in Local Schools*. Washington : National Institute of Education.
- Lipsey, M.W. 1988. « Practice and Malpractice in Evaluation Research ». *Evaluative Practice*, vol. 9, n° 4, p. 5-24.
- Madaus, G.F. et D.L. Stufflebeam. 2004. « Program Evaluation : A Historical Overview ». In *Evaluation Models : Viewpoints on Educational and Human Services Evaluation* (2^e éd.), sous la dir. de D.L. Stufflebeam, G.F. Madaus et Thomas Kellagan, p. 3-19. Boston : Kluwer Academic Publishers.
- McCarthy, T.A. 1979. *Communication and Evolution of Society : Jürgen Habermas*. Toronto : Beacon.
- _____. 1973. « A Theory of Communicative Competence ». *Philosophy of the Social Sciences*, vol. 3, p. 135-156.

- McDavid, J.C. 2001. « Program Evaluation In British Columbia in a Time of Transition : 1995-2000 ». *The Canadian Journal of Program Evaluation*, vol. Special Issues, p. 3-28.
- Owen, T. 1973. « Education Evaluation by Adversary Proceeding ». In *School Evaluation : The Politics and Process*, sous la dir. de E.R. House, Berkeley (Calif.) : McCutcheon.
- Patton, M.Q. 1978. *Utilization-Focused Evaluation*. Beverly Hills (Calif.) : Sage Publications.
- Patton, M. Q. 2002. *Qualitative Research Evaluation Methods*. 3rd edition. Thousand Oaks: Sage Publications.
- Preskill, H. et R.T. Torres. 1998. « Evaluative Inquiry as Transformative Learning ». *Panel/Paper Presented at the American Evaluation Association Annual Conference*.
- Provus, M. 1971. *Discrepancy Evaluation*, Berkeley (Calif.) : McCutchan Publishing.
- Redding, P. 1989. « Habermas' Theory of Argumentation ». *Journal of Value Inquiry*, vol. 23, p. 15-32.
- Rossi, P.H., H. E. Freeman et S. Wright. 1979. *Evaluation : A Systematic Approach*. Beverly Hills (Calif.) : Sage Publications.
- Rossi, P. H., M.W. Lipsey et H.E. Freeman. 2004. *Evaluation : A Systematic Approach*, Thousand Oaks (Calif.) : Sage Publications.
- Scriven, M. 1967. « The Methodology of Evaluation ». In *Curriculum Evaluation. American Educational Research Association Monograph Series on Evaluation no. 1*, sous la dir. de R.E. Stake, Chigago : Rand McNally.
- _____. 1972. « The Methodology of Evaluation ». In *Evaluating Action Programs : Readings in Social Action and Evaluation*, sous la dir. de C.H. Weiss, p. 123-136, Boston : Allyn & Bacon.
- _____. 1980. *The Logic of Evaluation*. Inverness (Calif.) : Edgepress.
- _____. 1986. « New Frontiers of Evaluation ». *Evaluation Practice*, vol. 7, p. 7-44.
- _____. 1991. *Evaluation Thesaurus* (4^e éd.). Newbury Park (Calif.) : Sage Publications.
- _____. 1995. « The Logic of Evaluation and Evaluation Practice ». *New Direction for Evaluation*, n° 68, p. 49-70.

- Stake, R.E. 1972. « The Countenance of Educational Evaluation ». In *Evaluating Action Programs : Readings in Social Action and Education*, sous la dir. de C.H. Weiss, p. 31-51. Boston : Allyn & Bacon.
- _____. 2004. *Standards-Based and Responsive Evaluation*. Thousand Oaks (Calif.) : Sage Publications.
- Stanley, J.C. et K.D. Hopkins. 1972. *Educational and Psychological Measurement and Evaluation*, Englewood Cliffs (N. J.) : Prentice Hall.
- Stufflebeam, D.L. 1983. « The CIPP Model for Program Evaluation ». In *Evaluation Models : Viewpoints on Educational and Human Services Evaluation*, G. F. Madaus, M.S. Scriven et D.L. Stufflebeam (éd.), p. 117-141, Boston : Kluwer-Nijhoff.
- Suchman, E. 1967. *Evaluative Research : Principles and Practice in Public Service and Social Action Program*, New York : Sage Publications.
- Taylor, P.W. 1961. *Normative Discourse*. Englewood Cliffs (N. J.) : Prentice Hall.
- Tyler, R.W. 1942. « General on Evaluation ». *Journal of Educational Research*, vol. 35, p. 492-501.
- Toulmin, S.E. 1964. *The Uses of Argument*. New York : Cambridge University.
- Toulmin S.E., R.D. Rieke et A. Janik. 1984. *An Introduction to Reasoning*. New York : Macmillan.
- Van der Maren, J.-M. 1996. *Méthode de recherche pour l'éducation*. Montréal : Presses de l'Université de Montréal.
- Weiss, C.H.. 1972. *Evaluating Action Programs : Readings in Social Action and Education*. Boston : Allyn & Bacon.
- Wholey, J. 1981. « Using Evaluation to Improve Program Performance ». In *Evaluation Research and Practice : Comparative and International Perspectives*, R.A. Levine, M.A. Solomon et G.M. Hellstern, p. 92-106. Beverly Hills (Calif.) : Sage Publications.
- Wholey, J.S., H.P. Hatry et K.E. Newcomer. 2004. *Handbook of Practical Program Evaluation*. San Francisco : Jossey-Bass.
- Wolf, R.L. 1975. « Trial by Jury : A New Evaluation Method ». *Phi Delta Kappan*, vol. 17, n° 11, p. 5-8.

APPENDICE A
GRILLE D'ANALYSE

RÉFÉRENCE BIBLIOGRAPHIQUE						
Revue	<i>Revue canadienne en évaluation de programme</i>		<i>American Journal of Evaluation</i>		<i>Evaluation and Program Planning</i>	
Année	1994		1999		2004	

CRITÈRES	BARÈME D'ÉVALUATION				
	NON	EN PARTIE		OUI	
		FAIBLE	FORT		
La description du programme : L'auteur décrit le programme de façon à ce que l'on puisse le comprendre.	La description est absente.	La description est à peine suffisante.	La description est présente mais succincte.	La description est présente et suffisante.	
La problématique (contexte) : L'auteur présente les motifs ou raisons qui justifient la réalisation de l'évaluation.	La problématique est inexistante.	La problématique est énoncée sans justification.	La problématique est énoncée et sa justification est implicite.	La problématique et sa justification sont clairement présentées.	
La ou les questions : L'auteur formule la ou les questions auxquelles l'évaluation doit apporter des éléments de réponse. La question doit être en lien avec la problématique.	La question est absente.			La question est présente.	
	La question n'est pas liée à la problématique.	La question ne couvre que quelques éléments de la problématique.	La question couvre la plupart des éléments de la problématique.	La question est en lien avec la problématique.	
	La question n'est pas formulée clairement.			La question est formulée clairement.	

CRITÈRES	BARÈME D'ÉVALUATION					
	NON	EN PARTIE				OUI
		FAIBLE		FORT		
<p>L'objectif et/ou le type d'évaluation (pris comme synonyme puisqu'un type d'évaluation sous-entend l'objectif) : L'auteur énonce les objectifs de l'évaluation selon la typologie de Rossi, Lipsey et Freeman (2004). L'objectif doit être en lien avec la question ou la problématique (une évaluation peut avoir plus d'un objectif).</p>						
• Objectif/type 1 :	L'objectif d'évaluation est absent.					L'objectif d'évaluation est présent.
	L'objectif d'évaluation n'est pas lié à la question.	L'objectif d'évaluation est peu lié à la question.		L'objectif d'évaluation est très lié à la question.		L'objectif d'évaluation est en lien avec la question.
	L'objectif d'évaluation n'est pas clairement formulé.					L'objectif d'évaluation est clairement formulé.
• Objectif/type 2 :	L'objectif d'évaluation est absent.					L'objectif d'évaluation est présent.
	L'objectif d'évaluation n'est pas lié à la question.	L'objectif d'évaluation est peu lié à la question.		L'objectif d'évaluation est très lié à la question.		L'objectif est en lien avec la question.
	L'objectif d'évaluation n'est pas clairement formulé.					L'objectif d'évaluation est clairement formulé.

CRITÈRES	BARÈME D'ÉVALUATION			
	NON	EN PARTIE		OUI
		FAIBLE	FORT	
<p>L'objet d'évaluation : L'auteur identifie la dimension du programme sur laquelle va porter le jugement. L'objet doit être en lien avec l'objectif et formulé clairement. Si l'évaluation a plusieurs objectifs ou types d'évaluation, utilisez les lignes ci-dessous.</p>				
○ Objet 1 (Objectif/type 1)	La dimension sur laquelle porte le jugement n'est pas identifiée.			La dimension sur laquelle porte le jugement est identifiée.
	L'objet n'est pas lié à l'objectif d'évaluation.	L'objet est partiellement en lien avec l'objectif d'évaluation.		L'objet est en lien avec l'objectif d'évaluation.
○ Objet 2 (Objectif/type 2)	La dimension n'est pas identifiée.			La dimension est identifiée.
	L'objet n'est pas lié à l'objectif d'évaluation.	L'objet est partiellement en lien avec l'objectif d'évaluation.		L'objet est en lien avec l'objectif d'évaluation.

CRITÈRES	BARÈME D'ÉVALUATION					
	NON	EN PARTIE			OUI	
		FAIBLE		FORT		
<p>Les critères : L'auteur identifie les points de référence sur lesquels il a basé son jugement. Le choix des critères est justifié. Les critères sont formulés clairement.</p> <p>Si plus d'un objectif :</p>						
○ Par rapport à l'objectif/type 1 :	Les critères sont absents.					Les critères sont énoncés.
	Le choix des critères n'est pas justifié.	Le choix des critères est à peine justifié.		Le choix des critères est implicitement justifié.		Le choix des critères est clairement justifié.
	Les critères ne sont pas clairement formulés.					Les critères sont clairement formulés.
○ Par rapport à l'objectif/type 2 :	Les critères sont absents.					Les critères sont énoncés.
	Le choix des critères n'est pas justifié.	Le choix des critères est à peine justifié.		Le choix des critères est implicitement justifié.		Le choix des critères est clairement justifié.
	Les critères ne sont pas clairement formulés.					Les critères sont clairement formulés.

CRITÈRES	BARÈME D'ÉVALUATION				
	NON	EN PARTIE		OUI	
		FAIBLE	FORT		
<p>Les variables observables : L'auteur identifie clairement les différentes variables observables qui serviront à mesurer les critères.</p> <p>Si plus d'un objectif :</p>					
○ Par rapport à l'objectif/type 1 :	Les variables observables ne sont pas clairement associées aux critères.	Les variables observables sont à peine associées aux critères.	Les variables observables sont implicitement associées aux critères.	Les variables observables sont clairement associées aux critères.	
○ Par rapport à l'objectif/type 2 :	Les variables observables ne sont pas clairement associées aux critères.	Les variables observables sont à peine associées aux critères.	Les variables observables sont implicitement associées aux critères.	Les variables observables sont clairement associées aux critères.	
<p>Les standards : L'auteur identifie des standards pour chaque critère ou variable observable. Ils doivent être en lien avec les variables observables ou les critères ?</p> <p>(* = ne s'applique pas, dans le cas d'une étude où il n'est pas possible de mesurer les résultats)</p> <p>Si plus d'un objectif :</p>					
○ Par rapport à l'objectif/type 1 :	Il n'existe pas de standards.		% des standards énoncés se rattachent aux critères.	Un standard est rattaché à chaque critère.	
○ Par rapport à l'objectif/type 2 :	Il n'existe pas de standards.		% des standards énoncés se rattachent aux critères.	Un standard est rattaché à chaque critère.	

CRITÈRES	BARÈME D'ÉVALUATION					
	NON	EN PARTIE		OUI		
		FAIBLE	FORT			
<p>Le jugement : L'auteur démontre que son jugement est établi à partir des écarts entre les critères et les standards.</p>	Le jugement n'est basé sur aucun fondement.			Le jugement émis s'appuie sur % des critères et standards.	Le jugement émis s'appuie sur les critères et standards.	
<p>Le jugement : L'auteur démontre que l'évaluation a répondu à la question d'évaluation.</p>	Le jugement n'est pas en lien avec la question, les objectifs.			Le jugement émis est en lien avec % des questions et objectifs.	Le jugement émis couvre l'ensemble des intentions de l'évaluation.	
<p>Autres variables :</p> <ul style="list-style-type: none"> • L'auteur présente le contexte théorique sur lequel il a fondé ses critères (présence d'un cadre théorique de référence). 	L'évaluation se fonde sur un contexte athéorique.				L'évaluation se fonde sur un contexte théorique.	
<ul style="list-style-type: none"> • L'auteur mentionne le recours à un modèle d'évaluation reconnu dans les écrits et il en justifie l'utilisation. 	Aucun modèle n'a été utilisé.				Le modèle est clairement présenté.	
<ul style="list-style-type: none"> ○ Si oui, lequel (modèle retenu)? <ul style="list-style-type: none"> ○ Axé sur la méthode ○ Axé sur le jugement ○ Axé sur la prise de décision 						

APPENDICE B
COEFFICIENT D'AGRÉMENT INTERJUGE

	Entente	Non-entente	Total	Po	Pc	Coefficient
Article 1	17	2	19	0,89	0,11	0,88
Article 2	32	7	39	0,82	0,18	0,78
Article 3	16	3	19	0,84	0,16	0,81
Article 4	12	5	17	0,71	0,29	0,58
Article 5	15	2	17	0,88	0,12	0,87
Article 6	16	1	17	0,94	0,06	0,94
Article 7	14	1	15	0,93	0,07	0,93
Article 8	14	3	17	0,82	0,18	0,79
Article 9	14	1	15	0,93	0,07	0,93
Article 10	15	2	17	0,88	0,12	0,87
Article 11	15	-	15	1,00	0,00	1,00
Article 12	16	1	17	0,94	0,06	0,94
Article 13	14	3	17	0,82	0,18	0,79
Article 14	14	3	17	0,82	0,18	0,79
Article 15	16	1	17	0,94	0,06	0,94
Coefficient global						0,85
Légende :						
Po : proportion des jugements sur lesquels il y a entente						
Pc : proportion des jugements sur lesquels il n'y a pas entente						
Calcul du coefficient : $Po - Pc / 1 - Pc$						

APPENDICE C COMPILATION DES RÉSULTATS

ID	Desc. prog.	Prob.	Question			Objectif			Objet		Critères			Var. obs.	Stan- dards	Jugement		Jug. Obj.	Cadre théoric	Modèle	Type
			Présente	En lien	Claire	Présent	En lien	Clair	Présent	En lien	Présents	Justifiés	Clairs			Critère	Std.				
EPP_04_1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	Impact
EPP_04_2	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	Coût/bén
EPP_04_3	1	1	2	1	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
EPP_04_4	1	2	4			1	2	1	1	1	1	2	1	2	3	1	3	1	1	2	Impl.
EPP_04_5	1	2	2	2	2	1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Coût/bén
EPP_04_6	2	2	2	2	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
EPP_04_7	2	2	2	1	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
EPP_04_8	1	2	2	2	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
EPP_04_9	1	1	2	1	2	1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Impact
EPP_04_10	2	2	2	2	2	1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Impact
EPP_04_11	1	2	2	1	2	1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Pert.
EPP_04_12	3	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Concept.
EPP_04_13	1	2	4			1	1	2	1	1	1	2	1	1	1	1	1	1	1	2	Impact

ID	Desc. prog.	Prob.	Question			Objectif			Objet		Critères			Var. obs.	Standards	Jugement		Jug. Obj.	Cadre théoric	Modèle	Type
			Présente	En lien	Clair	Présent	En lien	Clair	Présent	En lien	Présents	Justifiés	Clairs			Critère	Std.				
EPP_04_14	1	2	2	1	2	1	1	1	1	1	1	2	1	2	1	1	1	1	1	2	Coût/bén.
EPP_04_15	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_04_16	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impl.
EPP_04_17	2	2	4			1	2	2	1	1	1	2	2	2	3	1	3	1	1	2	Impact
EPP_04_18	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_04_19	2	2	2	1	2	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Pert.
CJE_04_1	2	2	4			1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Concept.
CJE_04_2	1	1	2	1	2	1	1	1	1	1	1	2	1	1	3	1	3	1	1	2	Impact
CJE_04_3	1	1	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
CJE_04_4	1	2	2	1	2	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Impl.
AJE_04_1	1	2	2	1	2	1	1	1	1	1	1	2	2	3	3	1	3	1	1	2	Impl.
AJE_04_2	1	2	4			1	1	1	1	1	1	1	1	2	2	1	2	1	1	2	Impact
AJE_04_3	2	2	2	2	2	1	2	2	1	1	1	3	2	3	3	1	3	1	1	2	Impact
AJE_04_4	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
AJE_04_5	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
AJE_04_6	2	2	4			1	1	2	1	1	1	2	1	2	2	1	2	1	1	2	Impact

ID	Desc. prog.	Prob.	Question			Objectif			Objet		Critères			Var. obs.	Standards	Jugement		Jug. Obj.	Cadre théoric	Modèle	Type
			Présente	En lien	Claire	Présent	En lien	Clair	Présent	En lien	Présents	Justifiés	Clairs			Critère	Std.				
AJE_04_7	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_99_1	1	2	2	1	2	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Pert.
EPP_99_2	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	Impact
EPP_99_3	1	2	2	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	Impact
EPP_99_4	1	1	2	1	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	2	Impact
EPP_99_5	1	2	2	1	2	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Pert.
EPP_99_6	1	2	1	1	1	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_99_7	2	2	2	1	2	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Impact
EPP_99_8	1	1	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_99_9	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	Impact
EPP_99_10	1	2	2	1	2	1	1	1	1	1	1	2	2	1	2	1	2	1	1	2	Impact
EPP_99_11	1	2	1	1	1	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Pert.
EPP_99_12	1	1	1	1	1	1	1	1	1	1	1	2	1	2	2	1	2	1	1	1	Impact
EPP_99_13	1	2	1	1	1	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_99_14	1	2	1	1	1	1	1	1	1	1	1	2	2	4	3	1	3	1	1	2	Impl.
EPP_99_15	1	2	1	1	1	1	1	1	1	1	1	2	1	1	2	1	2	1	1	1	Impact

ID	Desc. prog.	Prob.	Question			Objectif			Objet		Critères			Var. obs.	Standards	Jugement		Jug. Obj.	Cadre théoric	Modèle	Type
			Présente	En lien	Claire	Présent	En lien	Clair	Présent	En lien	Présents	Justifiés	Clairs			Critère	Std.				
EPP_99_16	1	2	2	1	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
CJE_99_1	1	3	4			1	1	1	1	1	1	2	1	3	2	1	2	1	2	2	Impact
CJE_99_2	2	2	2	1	2	1	1	1	1	1	1	2	2	3	3	1	3	1	1	2	Pert.
CJE_99_3	1	1	2	1	2	1	1	1	1	1	1	1	1	1	3	1	3	1	1	1	Impl.
CJE_99_4	1	2	1	1	1	1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Pert.
AJE_99_1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	Impact
AJE_99_2	2	3	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	Impact
AJE_99_3	2	3	1	1	1	1	1	1	1	1	1	3	2	4	2	1	2	1	2	2	Impact
EPP_94_1	3	2	4			1	2	1	1	1	1	2	1	2	3	1	3	1	1	2	Impl.
EPP_94_2	1	3	2	1	2	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	Impact
EPP_94_3	2	2	1	1	1	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_94_4	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
EPP_94_5	2	3	4			1	1	1	1	1	1	2	1	1	3	1	3	1	2	2	Concept.
EPP_94_6	1	1	2	1	2	1	1	1	1	1	1	2	1	4	1	1	1	1	1	2	Coût/bén.
EPP_94_7	2	2	4			1	1	1	1	1	1	2	1	2	3	1	3	1	1	2	Impact
EPP_94_8	1	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact

ID	Desc. prog.	Prob.	Question			Objectif			Objet		Critères			Var. obs.	Standards	Jugement		Jug. Obj.	Cadre théoric	Modèle	Type
			Présente	En lien	Clair	Présent	En lien	Clair	Présent	En lien	Présents	Justifiés	Clairs			Critère	Std.				
EPP_94_9	2	2	2	1	2	1	1	2	1	1	1	2	1	2	2	1	2	1	1	2	Coût/bén.
EPP_94_10	2	2	2	1	2	1	1	2	1	1	1	2	2	3	3	1	3	1	1	2	Pert.
EPP_94_11	2	2	4			1	1	2	1	1	1	2	2	3	3	1	3	1	1	2	Impact
EPP_94_12	1	2	2	1	2	1	1	1	1	1	1	2	2	2	2	1	2	1	1	2	Impact
CJE_94_1	1	2	4			1	1	1	1	1	1	2	1	1	3	1	3	1	1	1	Impl.
CJE_94_2	2	2	2	1	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	2	Impact
AJE_94_1	1	2	2	1	2	1	1	2	1	1	1	2	2	2	2	1	2	1	1	2	Impact
AJE_94_2	2	2	4			1	1	1	1	1	1	3	2	2	3	1	3	1	1	2	Impact

1 = Oui

2 = En partie – fort

3 = En partie – faible

4 = Non

Pert. = Pertinence

Impl. = Implantation

Concept. = Conceptualisation

Coût/bén. = Coût-bénéfice