

# Comprendre les pratiques évaluatives afin de les améliorer : la modélisation du processus spécifique à l'évaluation de programme

Marthe Hurteau, Guy Lachapelle et Sylvain Houle

Volume 29, numéro 3, 2006

La relation entre la théorie et la pratique en évaluation de programme : dialogue ET monologue

URI : <https://id.erudit.org/iderudit/1086392ar>

DOI : <https://doi.org/10.7202/1086392ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Hurteau, M., Lachapelle, G. & Houle, S. (2006). Comprendre les pratiques évaluatives afin de les améliorer : la modélisation du processus spécifique à l'évaluation de programme. *Mesure et évaluation en éducation*, 29(3), 27-44. <https://doi.org/10.7202/1086392ar>

Résumé de l'article

L'évaluation de programme est régulièrement remise en question quant à sa fonction première, soit sa capacité à poser un jugement éclairé sur les programmes à l'étude et à dégager des recommandations pertinentes. À ce jour, certains auteurs ont émis des hypothèses d'explication du phénomène sans pour autant les appuyer par une démarche scientifique. La présente recherche effectue une percée dans ce sens en se fixant comme objectif de valider ces pistes d'explication. Pour y parvenir, elle doit disposer d'une représentation opérationnelle du processus spécifique à l'évaluation de programme afin de décrire les pratiques. Hurteau et Houle (2005) proposent une telle représentation – modélisation – qui s'appuie sur différentes contributions théoriques, dont celle de Scriven (1980) qui a introduit le concept de « logique de l'évaluation » et celles d'autres auteurs (Hurteau, 1991 ; Fournier, 1995 ; Stake, 2004). La présente étude, qui constitue la première phase de la recherche, vise à présenter la modélisation ainsi que les résultats de sa validation.

## **Comprendre les pratiques évaluatives afin de les améliorer : la modélisation du processus spécifique à l'évaluation de programme**

**Marthe Hurteau**

*Université du Québec à Montréal*

**Guy Lachapelle**

*Université Concordia*

**Sylvain Houle**

*Université du Québec à Montréal*

**MOTS CLÉS :** Évaluation de programme; «logique de l'évaluation»; pratiques en évaluation de programme

*L'évaluation de programme est régulièrement remise en question quant à sa fonction première, soit sa capacité à poser un jugement éclairé sur les programmes à l'étude et à dégager des recommandations pertinentes. À ce jour, certains auteurs ont émis des hypothèses d'explication du phénomène sans pour autant les appuyer par une démarche scientifique. La présente recherche effectue une percée dans ce sens en se fixant comme objectif de valider ces pistes d'explication. Pour y parvenir, elle doit disposer d'une représentation opérationnelle du processus spécifique à l'évaluation de programme afin de décrire les pratiques. Hurteau et Houle (2005) proposent une telle représentation – modélisation – qui s'appuie sur différentes contributions théoriques, dont celle de Scriven (1980) qui a introduit le concept de «logique de l'évaluation» et celles d'autres auteurs (Hurteau, 1991; Fournier, 1995; Stake, 2004). La présente étude, qui constitue la première phase de la recherche, vise à présenter la modélisation ainsi que les résultats de sa validation.*

---

Note – Les auteurs de cet article tiennent à souligner la contribution de Stéphanie Mongiat, étudiante à la maîtrise en éducation, UQAM. Toute correspondance peut être envoyée par courriel à l'adresse suivante: [hurteau.marthe@uqam.ca].

KEY WORDS: Program evaluation; “logic of evaluation”; practices of program evaluation

*It has become common-place for clients to question program evaluation's primary function: its ability to produce judgments that lead to relevant recommendations. Some well-known authors have formulated hypotheses in an attempt to explain this situation; however, these hypotheses have not been supported by empirical research as yet. The present research carries out a breakthrough in this direction by having as objective to validate these explanations. For that purpose, it must have an operational representation of the process specific to program evaluation in order to describe the practice. Hurteau and Houle (2005) propose such a representation – modeling – which is based on various theoretical contributions, of which that of Scriven (1980) which introduced the concept of “logic of evaluation” and those of other authors (Hurteau, 1991; Fournier, 1995; Stake, 2004). The present study presents the modeling as well as the results of its validation.*

PALAVRAS — CHAVE: Avaliação de programas; «lógica de avaliação»; práticas em avaliação de programas

*A avaliação de programas é, habitualmente, questionada quanto à sua função principal, ou seja, a sua capacidade de produzir um juízo claro sobre os programas em estudo e de formular recomendações pertinentes. Actualmente, alguns autores emitiram hipóteses de explicação do fenómeno sem, para tal, as sustentar num procedimento científico. A presente investigação faz uma abertura neste sentido, tendo como objectivo validar estas pistas de explicação. Para o atingir, tem de dispor de uma representação operacional do processo específico de avaliação de programas a fim de descrever as práticas. Hurteau e Houle (2005) propõem esta representação – modelização – que se apoia em diferentes contributos teóricos, nomeadamente o de Scriven (1980) que introduziu o conceito de “lógica de avaliação” e os de outros autores (Hurteau, 1991; Fournier, 1995; Stake, 2004). O presente estudo, que constitui a primeira fase da investigação, propõe-se apresentar a modelização, assim como os resultados da sua validação.*

## Introduction

L'évaluation de programme a acquis ses lettres de noblesse grâce aux nombreux développements concernant autant ses fondements théoriques que sa pratique. Dans les faits, ces derniers ne semblent cependant pas s'arrimer de façon suffisamment harmonisée pour que cette activité soit en mesure de se distinguer des autres domaines qu'elle côtoie (King, 2003) et d'être reconnue comme une discipline (Dubois & Marceau, 2005). Ces constats amènent d'ailleurs plusieurs auteurs et chercheurs, pour ne mentionner que Scriven

(2001), Sanders (2001) et Patton (2001), à se préoccuper du sort que l'avenir lui réserve. En effet, l'évaluation de programme fait l'objet de nombreuses et régulières remises en question quant à sa capacité à remplir sa fonction première, c'est-à-dire poser un jugement éclairé sur les programmes à l'étude et dégager des recommandations pertinentes. Ce constat est d'ailleurs appuyé par certaines études, pour ne citer que celles réalisées par Virtanen et Uusikylä (2002), le Secrétariat du Conseil du Trésor du Canada (2004) et Toulemonde (2005). Ainsi, pour Virtanen et Uusikylä (2002), la pratique actuelle de l'évaluation de programme ne permet pas de porter un jugement qui prend en considération la complexité des programmes. L'étude du Secrétariat du Conseil du Trésor du Canada (2004), qui porte sur l'analyse de plus d'une centaine de rapports d'évaluation, conclut que 32 % d'entre eux présentent un raisonnement inadéquat pour appuyer leur jugement et que dans 50 % des cas, les jugements ne sont pas appuyés par une information pertinente. Finalement, Toulemonde (2005) conclut que les preneurs de décision utilisent très peu les résultats issus de démarches évaluatives parce qu'ils ne permettent pas de répondre aux questions posées.

Guba (1972) a été un des premiers auteurs à relever et à documenter ce malaise récurrent. Dans son article intitulé *The Failure of Educational Evaluation*, il a émis l'hypothèse que cette situation peut être en grande partie attribuable au fait que les évaluateurs éprouvent de la difficulté à bien cerner et définir les objets d'évaluation («*evaluand*») ainsi que les critères requis pour effectuer une évaluation efficace. Scriven (1995) souscrit à cette analyse et renchérit en avançant que l'information qui émane d'un grand nombre d'évaluations de programme se révèle la plupart du temps, certes étoffée, mais inadéquate pour appuyer le jugement formulé. Ces deux auteurs émettent ainsi la possibilité que les difficultés auxquelles se confronte l'évaluation de programme pourraient s'expliquer par un manque de cohérence quant au processus logique sous-jacent à toute démarche évaluative. Considérant que ces constats n'ont jamais fait l'objet d'une vérification, et ce, malgré le fait qu'ils constituent, à ce jour, les seules explications documentées, la présente recherche vise à établir dans quelle mesure les incohérences soulevées par Scriven et Guba peuvent effectivement s'avérer une explication pertinente et adéquate. Ces visées s'inscrivent dans la ligne de pensée de certains auteurs, pour ne citer que Christie (2005), Henry & Mark (2003) et King (2003), qui sont d'avis que les difficultés pourraient être surmontées si des recherches s'attardaient à décrire et à comprendre la pratique évaluative afin de lui fournir les fondements théoriques requis. À notre connaissance, il existe peu de publi-

cations scientifiques qui rapportent des recherches ayant pour objet d'étude la démarche évaluative et celles qui existent ont peu contribué à la solution du problème parce qu'elles n'établissent pas de ponts significatifs entre les conceptions théoriques et la pratique (Chandler, 2001 ; Christie, 2003b ; Datta, 2003 ; Dubois & Marceau, 2005). Les recherches de Christie (2003a) constituent un effort dans ce sens mais elles s'appuient sur ce que les évaluateurs «perçoivent faire» et non pas sur ce qu'ils «font dans les faits».

Si la pertinence de la présente recherche est assurée, il n'en demeure pas moins qu'elle se heurte à un problème non négligeable puisqu'elle désire décrire les pratiques évaluatives, afin d'éventuellement conclure si, effectivement, on peut confirmer les incohérences soulevées par Scriven et Guba, et qu'il existe peu d'écrits lui permettant de s'appuyer sur une définition ou une représentation opérationnelle du processus logique de l'évaluation. C'est dans cette perspective qu'il est devenu évident que la première phase de la présente recherche doit élaborer et valider une modélisation – une représentation – du processus spécifique à l'évaluation de programme, afin d'être en mesure, éventuellement, de décrire les pratiques évaluatives. Ces opérations constituent le propos du présent article.

## État de la question

Certains concepts théoriques de base seront maintenant développés afin de permettre au lecteur de saisir les notions sous-jacentes qui ont contribué à l'élaboration de la modélisation du processus spécifique à l'évaluation de programme.

### *L'évaluation de programme*

Elle se définit comme une activité scientifique visant à établir la valeur ou le mérite d'un programme ou d'un de ses éléments constitutifs (House, 1980), et ce, au moyen de déclarations et de conclusions qui sont précises, justifiables, acceptables, légitimes ; bref, qui sont valides (Fournier & Smith, 1993). Indépendamment de la position épistémologique choisie, la définition de l'évaluation de programme ne pourrait se défaire de sa fonction d'émettre un jugement concernant la valeur, le mérite, la signifiante ou les qualités sur le programme à l'étude puisque le terme «évaluer» lui est conceptuellement et pratiquement attaché à l'intérieur du terme «évaluation de programme». Elle implique également une cueillette systématique de données sur l'objet à l'étude, soit le programme, afin que la conclusion de cette démarche soit argumentée par une

information dite suffisante pour être acceptée du client de l'évaluation. Les conclusions évaluatives pourront alors servir à la prise de décision concernant la programmation future du programme (Patton, 2001 ; Stake, 2004 ; Weiss, 1997). La notion de «jugement», que Scriven (1980) associe à la «logique de l'évaluation», s'avère cruciale puisqu'elle permet de distinguer l'évaluation des autres activités.

### ***La théorie de l'évaluation de programme***

Différentes théories ont vu le jour pour encadrer la pratique de l'évaluation de programme. Chacune à leur manière, elles précisent les différents éléments qui la définissent et qui la caractérisent. Des auteurs ont regroupé l'ensemble de ces théories afin d'en dégager les éléments essentiels. C'est ainsi que sont apparues les deux «métathéories» actuellement reconnues (Melvin, 2006), soit celle de Shadish, Cook et Leviton (1991) et celle de Alkin (2004). La première synthétise la pratique de l'évaluation de programme au moyen des cinq composantes suivantes :

1. *la perspective des programmes sociaux ;*
2. *la construction du savoir ;*
3. *le jugement de valeur ;*
4. *l'utilisation des résultats ;*
5. *la pratique.*

La seconde (Alkin, 2004), quant à elle, propose de regrouper les modèles d'évaluation de programme selon trois axes, à savoir :

1. l'utilisation (*use*),
2. la méthodologie (*method*),
3. le jugement de valeur (*value*).

Cette nomenclature met en évidence la contribution ainsi que la spécificité de chacun des axes. Elle fait aussi ressortir les particularités de chacun des modèles d'évaluation en fonction de caractéristiques telles que le niveau et la nature de l'engagement des personnes visées par le programme (*stakeholders*) sur le plan de la démarche évaluative ; l'importance relative accordée à l'utilisation des résultats ; des considérations plus reliées aux fondements – la théorie du programme et la justice sociale ; finalement, les méthodes utilisées (Alkin & Christie, 2005).

Il ressort de cette brève présentation de l'évaluation de programme ainsi que des «métathéories» qui constituent actuellement ses fondements que, non seulement le jugement est-il spécifique à cette activité, mais qu'il est aussi intégré au sein des deux métathéories, confirmant ainsi sa contribution significative.

### ***Le jugement au sein de l'évaluation de programme***

Selon Shadish, Cook et Leviton (1991), la «logique de l'évaluation» introduite par Scriven (1980) constitue une autre forme de «métathéorie» qui place d'ailleurs le jugement au cœur de la démarche. L'auteur cerne les quatre opérations qui définissent le processus logique propre à l'évaluation de programme, soit :

1. établir les critères (les éléments de l'objet qui seront évalués) afin de déterminer les attributs de l'objet d'évaluation sur lesquels reposera le jugement (Mathison, 2004) ;
2. élaborer les standards afin d'établir le niveau de performance souhaité qui constitue en quelque sorte le point de coupure entre ce qui constitue un jugement favorable ou non ;
3. mesurer la performance et la mettre en relation avec les standards, ce qui constitue la cueillette des données et leur analyse ;
4. synthétiser et intégrer les données afin d'être en mesure de porter un jugement.

Ce cadre conceptuel s'appuie sur la théorie de la valeur des jugements (Hare, 1972) qui stipule qu'un jugement trouve sa valeur en s'appuyant sur la validité de ses inférences. Plusieurs auteurs accordent une crédibilité à la «logique de l'évaluation», dont Rossi, Lipsey et Freeman (2004) qui l'incluent dans leur chapitre traitant de l'établissement de ce que constitue une bonne question d'évaluation, élément crucial dans la planification de toute évaluation qui se veut cohérente.

Plusieurs auteurs ont poursuivi la réflexion initiale de Scriven (1980) sur ce qui peut constituer l'acte spécifique d'évaluer, pour ne citer que Hurteau (1991), Fournier (1995) et Stake (2004). Ces différentes contributions présentent l'avantage de fournir une perspective opérationnelle à cette «métathéorie» que constitue la «logique de l'évaluation».

Considérant que Guba (1972) et Scriven (1995) établissent une relation entre un manque de cohérence à l'intérieur des éléments associés à la «logique de l'évaluation» et les nombreuses remises en question auxquelles se confronte l'évaluation de programme, la présente recherche vise à valider les réflexions

émises par Guba et Scriven en se posant la question : dans quelle mesure la « logique de l'évaluation » est-elle effectivement intégrée à la pratique évaluative et dans quelle mesure peut-elle être associée aux difficultés qu'elle rencontre ? Plus spécifiquement, la recherche poursuit les objectifs suivants :

- établir dans quelle mesure la « logique de l'évaluation » dépasse le statut de métathéorie pour être opérationnelle et décrire la praxie ;
- établir dans quelle mesure elle permet de discriminer les évaluations « appréciées » par les publics cibles de celles qui ne le sont pas afin de poser les premiers jalons de ce qui pourrait constituer une explication de l'« échec de l'évaluation ».

Ultimement, non seulement les résultats qui se dégageront des différentes activités pourront-ils offrir des pistes de solution aux problèmes auxquels est confrontée l'évaluation de programme mais ils contribueront aussi au développement de la théorie de l'évaluation qui encadrerait la pratique. Le présent article relate la première phase de la recherche, qui a consisté à développer et valider une modélisation du processus spécifique à l'évaluation de programme afin qu'elle puisse être utilisée par la suite pour décrire la pratique évaluative.

## Méthode

Deux opérations ont été retenues pour développer une modélisation du processus spécifique à l'évaluation de programme et pour la valider. La première a consisté à réaliser une anasynthèse afin de déterminer les éléments et la forme que pourrait prendre cette modélisation (Landry, 2006 ; Landry & Auger, 2003 ; Legendre, 1993 ; Silvern, 1972) et, ainsi, rendre opérationnel le concept de « logique de l'évaluation ». La seconde opération a consisté à la valider.

### *Première opération : l'élaboration de la modélisation*

L'anasynthèse, qui a été retenue dans le contexte actuel parce qu'elle offre les balises requises pour élaborer la modélisation du processus spécifique à l'évaluation de programme, s'intéresse plus spécifiquement aux différentes contributions théoriques traitant du processus logique qui sous-tend toute démarche évaluative, cette thématique étant à l'origine de la présente réflexion. Il est important de mentionner que les différentes opérations de schématisation de la modélisation ainsi que sa validation de contenu auprès d'experts ont respecté le critère de « saturation ».

Tel que mentionné en introduction, Hurteau (1991), Fournier (1995) et Stake (2004) ont assuré le développement du concept de «logique de l'évaluation» en poursuivant la réflexion amorcée par Scriven (1980). En effet, Hurteau (1991) introduit et distingue deux types de choix sous-jacents à toute démarche évaluative, soit les «choix stratégiques» et les «choix méthodologiques». Les premiers s'inscrivent dans la ligne de pensée de la «logique de l'évaluation» et font référence à l'objet de l'évaluation, aux critères retenus pour qualifier l'objet ainsi qu'aux standards sélectionnés. Les seconds incluent le choix du modèle d'évaluation ainsi que des considérations méthodologiques propres à tout type d'enquête. Fournier (1995), pour sa part, introduit les concepts de «logique générale» et de «logique de travail». Alors que la première reprend intégralement les termes de la «logique de l'évaluation», la seconde, pour sa part, offre quatre nouveaux éléments qui permettent l'application des opérations dans un contexte d'évaluation de programme, soit : le problème, l'objet d'étude, la question et les conclusions. Ces éléments confèrent ainsi un caractère plus tactique à la démarche. Finalement, Stake (2004) propose deux types de pensées évaluatives qui visent à structurer une démarche évaluative, soit la «pensée critériée» et la «pensée interprétative». La première fait référence aux fondements de la démarche évaluative, à savoir l'établissement des objets, critères et standards, et se rapproche par le fait même de la «logique de l'évaluation» (Scriven 1980) ainsi que des «choix stratégiques» (Hurteau 1991). La seconde, quant à elle, se préoccupe davantage de la nature des relations avec le client et de la façon dont cette dernière va teinter l'établissement des éléments liés à la «pensée critériée». Dans ce sens, elle assure le pont entre les aspects stratégiques et méthodologiques de la démarche évaluative.

La convergence des points de vue semble assez évidente, ce qui leur confère, par le fait même, une certaine validité. En effet, les différents auteurs ont cerné la spécificité du processus d'évaluation de programme, que ce soit par les termes de «logique de l'évaluation», de «choix stratégiques», de «logique générale» ou de «pensée critériée». De plus, tous établissent une distinction entre les composantes de la démarche évaluative et les considérations méthodologiques, certes requises pour mener à terme l'évaluation mais non spécifiques à cette activité.

La modélisation du processus spécifique à l'évaluation de programme qui découle de l'anasythèse est illustrée à la figure 1.

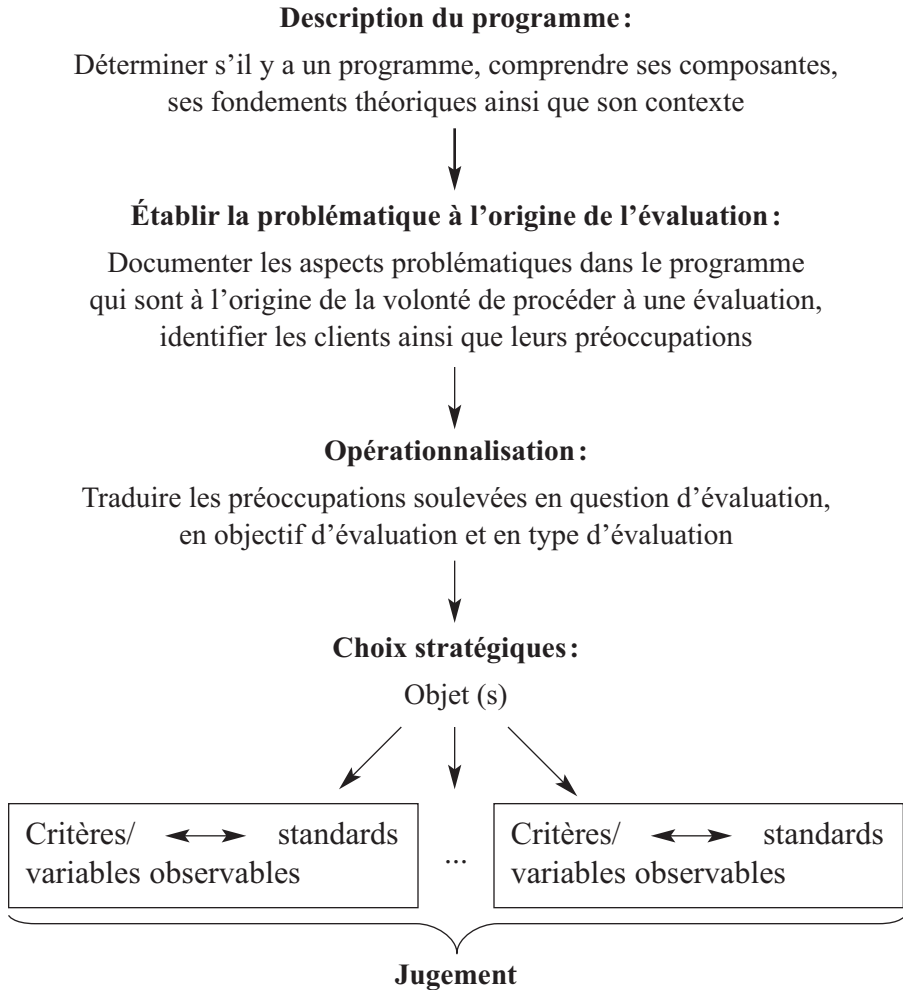


Figure 1 *Modélisation du processus spécifique à l'évaluation de programme: éléments et interrelations.*

Plus précisément, le «programme à évaluer» fait référence aux aspects qu'il est nécessaire de documenter afin de s'assurer qu'il y a bien un programme qui pourra faire l'objet d'une évaluation. La «problématique» constitue une étape cruciale à la démarche puisqu'elle inclut l'élément déclencheur de l'évaluation et les préoccupations des clientèles visées. L'«opérationnalisation» permet de traduire les besoins présents à l'intérieur de la problématique en langage propre à l'évaluation de programme. Les «choix stratégiques» constituent la pierre angulaire de la démarche évaluative

puisque'ils enclenchent la détermination des critères ainsi que des standards qui sont essentiels pour porter un jugement étoffé sur le programme. Finalement, le «jugement» repose sur l'analyse des écarts entre la mesure des variables observables et les standards respectifs (Scriven, 1980). Même si l'illustration de ce processus apparaît linéaire, il est important de souligner que son application, elle, est **itérative**. De plus, elle implique nécessairement la prise en considération de l'instrumentation qui est requise pour réaliser la démarche évaluative, pour ne citer que les méthodes, les modèles d'évaluation de programme et les responsabilités de l'évaluateur définies dans les standards du Joint Committee on Standards for Educational Evaluation (1981) ainsi que les lignes directives en éthique.

### ***La seconde opération : la validation de la modélisation***

Pour établir la validation, il a été décidé de confronter la modélisation à des pratiques évaluatives jugées exemplaires. Les chercheurs postulent que si la modélisation s'observe au sein de pratiques exemplaires, il sera alors envisageable qu'elle puisse être considérée comme une représentation efficace du processus spécifique à l'évaluation de programme. Ce même type d'argument est d'ailleurs invoqué dans l'étude de Arens (2005). L'ensemble de cette démarche sera maintenant développé puisqu'il constitue l'essence du présent propos.

#### ***Échantillon***

L'étude porte sur l'analyse de 69 évaluations de programme ayant fait l'objet d'une publication dans trois revues scientifiques spécialisées (*Journal canadien en évaluation de programmes*, *American Journal of Evaluation*, *Evaluation and Program Planning*), ce qui leur confère un statut de «pratiques exemplaires». Leur sélection s'est effectuée à partir de critères préalablement définis :

- articles qui relatent exclusivement l'ensemble des phases d'une démarche d'évaluation de programme ;
- couvrant les dix dernières années,

en sélectionnant la population des articles concernés pour les années 1994, 1999 et 2004 plutôt que d'échantillonner au hasard les articles couvrant cette période. Ce choix méthodologique assurait ainsi que la période de temps était également couverte. Le tableau 1 présente la répartition des articles selon les revues et les années. Il est à noter que seulement 25 % des articles présents dans les revues scientifiques ont été retenus aux fins de notre analyse, les autres textes portant sur des sujets autres que la présentation d'une évaluation de programme.

Tableau 1  
*Répartition des articles analysés*

| <i>Articles</i> | <i>1994</i> | <i>1999</i> | <i>2004</i> | <i>Total</i> |
|-----------------|-------------|-------------|-------------|--------------|
| Retenus         | 16          | 23          | 30          | 69           |
| Total           | 77          | 95          | 89          | 261          |

### ***Information recueillie***

Les articles retenus ont fait l'objet d'une analyse de contenu au moyen d'une grille élaborée à cette fin et qui traduisait les éléments de la modélisation. Les indicateurs proposés définissaient de façon opérationnelle :

- la présence des éléments de la modélisation ;
- l'interrelation de ces éléments.

La grille permet de porter trois niveaux d'appréciation sur chacun des indicateurs, soit : absent, présent ou présent en partie. Si les deux premiers niveaux d'appréciation s'avèrent relativement aisés à poser, le troisième – présent en partie – peut se révéler plus arbitraire. C'est dans cette perspective que chaque niveau d'appréciation, pour chacun des indicateurs, s'accompagne d'une définition opérationnelle.

### ***Rigueur***

Dans un souci de rigueur (Neuendorf, 2001), la grille a été élaborée et validée sur le plan du contenu au moyen de deux stratégies. D'une part, la correspondance entre ses items et des considérations théoriques a été établie ; d'autre part, trois experts dans le domaine ont émis une appréciation positive et ils ont formulé des suggestions pour la bonifier. La stabilité des jugements, quant à elle, a été vérifiée au moyen de l'établissement du pourcentage d'accord entre des juges qui a été fixé à au moins 80 %. Cette démarche a été effectuée en appliquant la grille aux articles qui n'avaient pas été retenus aux fins de la présente étude et le pourcentage d'accord a été établi. L'analyse de contenu des articles comme telle n'a pas commencé avant que le pourcentage d'accord entre les juges s'avère supérieur à 80 %, et ce, de façon constante. Cette opération a d'ailleurs été régulièrement reprise en cours d'étude afin de s'assurer de la stabilité des appréciations portées. Plus spécifiquement, la démarche d'analyse de contenu s'est inspirée des cadres de référence offerts par Dillon (1984) et Neuendorf (2001). Finalement, compte tenu de l'importance que revêt cette démarche sur les prochaines étapes de la recherche, il a été décidé que la modélisation serait considérée valide à condition qu'elle se conforme

au standard suivant : une représentation des éléments et des interrelations de la modélisation qui s'observe dans au moins 80 % des cas. Une marge d'erreur ayant été établie à 5 %, le pourcentage demeure acceptable s'il oscille entre 76 % et 80 %.

## Résultats<sup>1</sup>

Il faut se rappeler que la première phase de la recherche consiste à développer et valider une modélisation du processus spécifique à l'évaluation de programme afin de l'utiliser, par la suite, pour décrire la pratique. Plus spécifiquement, l'étude vise à établir dans quelle mesure les éléments de la modélisation et leurs interrelations sont observés au sein de ce qui pourrait être considéré comme des pratiques exemplaires, 69 articles de revues scientifiques dans le cas présent. L'information contenue au tableau 2 permet de répondre en grande partie à cette question.

Tableau 2

### *Distribution des fréquences pour les éléments de la modélisation*

| <i>Éléments et interrelations<br/>de la modélisation</i> | <i>Présence</i>                |                  |            |
|----------------------------------------------------------|--------------------------------|------------------|------------|
|                                                          | <i>Fréquence (pourcentage)</i> |                  |            |
|                                                          | <i>Oui</i>                     | <i>En partie</i> | <i>Non</i> |
| <b>Description du programme</b>                          | 47 (68%)                       | 22 (32%)         |            |
| <b>Établissement de la problématique</b>                 | 14 (20%)                       | 55 (80%)         |            |
| <b>Opérationnalisation :</b>                             |                                |                  |            |
| - Question d'évaluation formulée                         | 12 (17%)                       | 44 (64%)         | 13 (19%)   |
| - Question en lien avec la problématique                 | 51 (74%)                       | 5 (7%)           | 13 (19%)   |
| - Objectif de l'évaluation formulé                       | 60 (87%)                       |                  | 9 (13%)    |
| - Objectif en lien avec la question                      | 65 (94%)                       |                  |            |
| - Type d'évaluation formulé                              | 60 (87%)                       | 4 (6%)           |            |
| <b>Choix stratégiques :</b>                              |                                |                  |            |
| - Objet(s) présenté(s)                                   | 69 (100%)                      |                  |            |
| - Critères annoncés                                      | 69 (100%)                      |                  |            |
| - Critères justifiés                                     | 7 (10%)                        | 62 (90%)         |            |
| - Standards annoncés                                     | 42 (61%)                       |                  | 27 (39%)   |
| <b>Jugement :</b>                                        |                                |                  |            |
| - Jugement couvre les objectifs d'évaluation             | 69 (100%)                      |                  |            |
| - Jugement basé sur des critères                         | 69 (100%)                      |                  |            |
| - Jugement basé sur des standards                        | 42 (61%)                       |                  | 27 (39%)   |

Plusieurs données intéressantes émergent de ces résultats. Ainsi :

- Mis à part la présentation de la problématique et de la question clairement formulée, la majorité des autres paramètres du processus sont effectivement présents dans plus de 80 % des cas. De plus, si la question n'est pas toujours clairement formulée, il n'en demeure pas moins qu'elle est en lien avec la problématique, ce qui constitue un élément important dans une perspective de cohérence du processus.
- Les jugements émis couvrent les objectifs annoncés et se réfèrent aux critères annoncés. Cependant, il est à noter que la présente étude n'a pas approfondi la rigueur de l'argumentation sur laquelle s'appuient les jugements.
- Fait intéressant, il est à noter que les standards sont absents dans un pourcentage assez élevé, soit 40 %. De plus, les évaluateurs y ont surtout recours dans le cadre d'évaluation des résultats en ayant recours à des schémas expérimentaux ou quasi-expérimentaux – standards comparatifs.
- Finalement, un des résultats les plus révélateurs qui émerge de cette étude demeure le fait que l'absence de standards s'accompagne souvent d'une difficulté à dépasser le stade de la simple description – des évidences – pour porter un jugement. Et dans le cas où les évaluateurs se risquent à émettre un jugement, ce dernier demeure plus souvent qu'autrement vague, du ressort des généralités. S'il est vrai que ce constat s'appuie sur des impressions recueillies en cours de route et non sur des données recueillies à partir de la grille, comme ce fut le cas pour les autres éléments de la modélisation, il n'en demeure par moins que l'étude de Arens (2005) arrive aux mêmes constatations, ce qui leur assure par le fait même une certaine crédibilité.

Ainsi, l'attribution d'un score de 2 au résultat «oui», de 1 au résultat «en partie» et de 0 au résultat «non», génère un score global de 1 544, comparativement au score absolu de 1 932 ( $14 \times 2 \times 69 = 1\,932$ ). Le pourcentage de présence des éléments et des interrelations de la modélisation étant de 80 %, donc le standard étant rencontré, la modélisation peut donc être confirmée dans son statut de représentation du processus spécifique à l'évaluation de programme.

## Forces et limites de l'étude

La présente étude revêt un certain intérêt par son effort à dépasser les spéculations traditionnelles concernant les pratiques évaluatives (*ce que les évaluateurs pensent faire* par rapport à *ce qu'ils font*) pour présenter des faits systématiques et documentés concernant les pratiques évaluatives qui s'avèrent des pratiques exemplaires dans le présent exercice. De plus, les faits qui se dégagent de cette démarche s'appuient sur un cadre théorique, lui-même documenté. Il est à noter que l'éventuel impact de l'absence de standard sur le jugement émis n'a pas été documenté avec la même rigueur que les autres aspects de l'étude et il a été présenté comme tel.

## Discussion

La présente étude a eu recours à un standard global et c'est dans ces conditions que la modélisation est confirmée. Par contre, il en aurait été autrement si un standard holistique avait été appliqué. En effet, celui-ci exige que chacun des éléments et chacune des interrelations de la modélisation se conforme au standard (80%) et cela n'aurait pas été le cas pour les standards qui ne se retrouvent que dans 60% des cas. Nous considérons que la conclusion à l'effet que la modélisation se trouve validée par la présente démarche est maintenue, parce que nous n'attribuons pas tant ce faible résultat à la pertinence des standards dans la modélisation comme au fait qu'il met en évidence la complexité de fixer des standards. Ce problème est d'ailleurs souligné par certains auteurs. Il n'en demeure pas moins que cette dernière observation peut constituer une piste d'explication intéressante concernant les problèmes cités en prémisses, à savoir la remise en question de la fonction première de l'évaluation de programme ainsi que le peu de crédibilité accordée aux résultats générés de la part des clients.

La réflexion concernant ce qui pourrait expliquer le problème de crédibilité des jugements a amené les chercheurs à approfondir la façon dont ils sont produits et argumentés. L'étude des travaux de Scriven (1995), Fournier (1995) et Toulmin (1993) et la prise en considération des récents travaux de Arens (2005) ont permis de préciser certains éléments de la modélisation, qui sont illustrés à la figure 2.

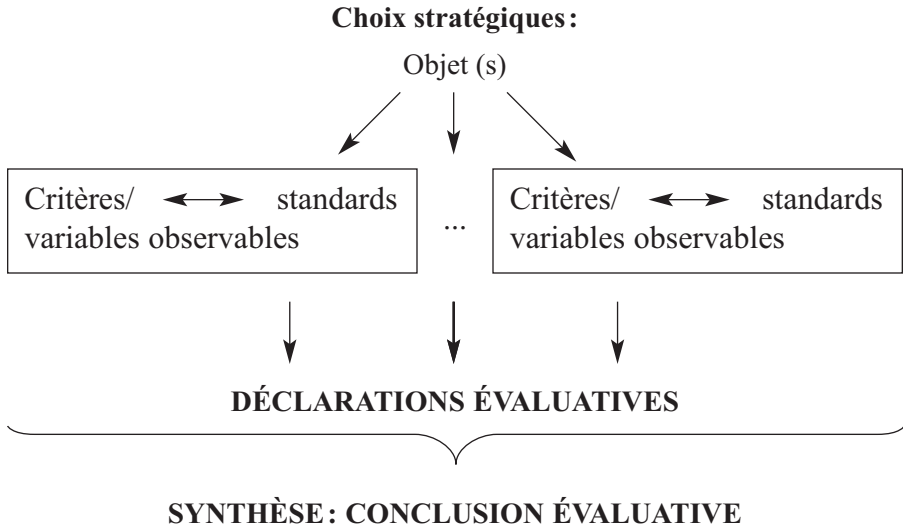


Figure 2 *Ajustement de la modélisation du processus spécifique à l'évaluation de programme.*

La modélisation intègre ainsi les récentes réflexions de Scriven (2006) qui remplace la notion de jugement par « synthèse » ou « conclusion évaluative ». Il la définit comme une forme de décision ou d'opinion qui s'appuie

- sur des faits pertinents ;
- sur la performance observée, elle-même établie à partir d'indicateurs de mérite ;
- sur la comparaison de performances établies et reconnues ou sur des standards validés ;
- sur l'impact de facteurs contextuels ainsi que d'un étalonnage.

Ainsi, les différentes déclarations concernant un aspect de l'objet d'évaluation sont synthétisées afin de produire une ou plusieurs déclarations plus globales qui, elles-mêmes, seront finalement transformées en conclusions évaluatives. S'il existe une variété d'approches pour générer des conclusions évaluatives, toutes se basent cependant sur un principe commun où une forme de pondération est appliquée aux différentes déclarations. La phase 2 du projet intégrera ces nouvelles considérations et les approfondira.

## Conclusion

Au cours des vingt dernières années, des efforts importants ont été investis sur le plan du développement de l'évaluation de programme. Cependant, elle se confronte toujours aux nombreuses et constantes remises en question qui proviennent principalement des clients. Il semblerait que le peu de recherches fondamentales sur les pratiques n'a pas réussi à dégager une théorie de l'évaluation de programme pouvant encadrer adéquatement la pratique. Les réflexions de Dubois et Marceau (2005) ainsi que de King (2003) vont dans le même sens. La présente étude constitue le premier jalon d'un effort pour cerner la nature des problèmes rencontrés afin d'apporter des correctifs pertinents et efficaces.

Il a été peu question de l'évaluation de programme appliquée au domaine de l'éducation, qui a connu au cours des cinquante dernières années un développement important, passant de l'évaluation du rendement académique des étudiants (Tyler) à des évaluations dont les enjeux sont de plus en plus complexes, et ce, dans la perspective de prendre des décisions lourdes de conséquences. Il faut dire que si les éventuelles retombées de la présente recherche seront significatives pour l'évaluation de programme en général, il n'en demeure pas moins qu'elles revêteront une pertinence tout aussi importante pour le domaine de l'évaluation de programme en éducation. Dans ce sens, l'amélioration de ses pratiques évaluatives ne peut faire autrement que de passer par l'amélioration des pratiques en général.

## NOTE

1. Les résultats ont été générés dans le cadre des travaux menant à l'obtention de la maîtrise de Julie Duval (Faculté des sciences de l'éducation, Université du Québec à Montréal), sous la direction de Marthe Hurteau.

## RÉFÉRENCES

- Alkin, M. C. (éd.) (2004). *Evaluation Roots. Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage Publications.
- Alkin, M. C., & Christie, C. A. (2005). Unraveling Theorists' Evaluation Reality. In M. A. Alkin & C. A. Christies (éds), *Theorists' Models in Action* (pp. 111-128). *New Directions for Evaluation*, 106. San Francisco: Jossey-Bass.
- Arens, S. A. (2005). *Understanding practice through exemplary evaluation cases*. Joint CES/AEA Conference, Toronto.
- Chandler, M. (2001). *How evaluators engage theory and philosophy in their practice*. Communication présentée à la conférence annuelle de l'American Evaluation Association, St. Louis, novembre 2001.

- Christie, C. A. (2003a). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory. *New Direction for Evaluation*, 97, 7-36. San Francisco: Jossey-Bass.
- Christie, C. A. (2003b). Understanding Evaluation Theory and its Role in Guiding Practice: Formal, Folk, and Otherwise. In C. Christina, *The Practice-Theory Relationship in Evaluation*, *New Directions for Evaluation*, 97, 91-93. San Francisco: Jossey-Bass.
- Christie, C. A. (éd.) (2005). Theorists' Models in Action. *New Direction for Evaluation*, 106. San Francisco: Jossey-Bass.
- Datta, L. E. (2003). Important questions, intriguing method, incomplete answers. In C. Christina, *The Practice-Theory Relationship in Evaluation*, *New Directions for Evaluation*, 97, 37-46. San Francisco: Jossey-Bass.
- Dillon, J. T. (1984). The classification of research questions. *Review of Educational Research*, 54(3), 327-361.
- Dubois, N., & Marceau, R. (2005). Un état des lieux théoriques de l'évaluation: une discipline à la remorque d'une révolution scientifique qui n'en finit pas. *Journal canadien en évaluation de programme*, 20(1), 1-36.
- Duval, J. (2005). *Étude descriptive des pratiques évaluatives: recours à la «logique de l'évaluation»*. Mémoire présenté dans le cadre de la maîtrise en éducation, Faculté des sciences de l'éducation, UQAM.
- Fournier, D. M. (1995). Establishing Evaluative Conclusions: A Distinction Between General and Working Logic. *New Direction for Evaluation*, 68, 15-32. San Francisco: Jossey-Bass.
- Fournier, D. M., & Smith, N. L. (1993). Clarifying the merits of argument in evaluation practice. *Evaluation and Program Planning*, 16(4), 315-323.
- Guba, E. G. (1972). The Failure of Educational Evaluation. In C. H. Weiss (éd.), *Evaluating Action Programs: Readings in Social Action and Education* (pp. 251-266). Boston: Allyn & Bacon Inc.
- Hare, R. M. (1972). What is a Value Judgment? In P. W. Taylor (éd.), *Problems of Moral Philosophy: Introduction to Ethics*. Encino: Dickenson.
- Henry, G. T., & Mark, M. M. (2003). The Challenge of Studying Evaluation Theory. *New Direction for Evaluation*, 97, 69-80.
- House, E. R. (1980). *Evaluating with validity*. Beverly Hills, CA: Sage.
- Hurteau, M. (1991). Strategic Choices in Programs Evaluation. In A. E. Love (éd.), *Sourcebook in Evaluation* (pp. 106-125). Ottawa: Les publications de la Société canadienne en évaluation.
- Hurteau, M., & Houle, S. (2005). *Identifying a Core Body of Knowledge for Evaluators*. Joint CES/AEA Conference, Toronto. Acte du congrès [<http://lg.evaluationcanada.ca>].
- Hurteau, M., Houle, S., & Duval, J. (2005). *L'étude des pratiques évaluatives: Un retour aux fondements de la discipline*. Joint CES/AEA Conference, Toronto.
- Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- King, J. A. (2003). The Challenge of Studying Evaluation Theory. In C. A. Christie, *The Practice-Theory Relationship in Evaluation*, *New Direction for Evaluation*, 97, 57-67. San Francisco: Jossey-Bass.

- Landry, N. (2006). *Vers une classification du domaine perceptuel en éducation préprimaire. Proposition d'un construit théorique*. Thèse de doctorat inédite, Université du Québec à Montréal.
- Landry, N., & Auger, R. (2003). *Stratégies de clarification conceptuelle*. Document RT\_LABDORM300303. Laboratoire d'analyse de données et de formation en mesure et évaluation.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation* (2<sup>e</sup> édition). Montréal: Guérin.
- Mathison, S., et al. (2004). *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage Publications.
- Melvin, M. (2006). *Evaluation theory Past, Present, and Future*. Présentation donnée dans le cadre du congrès annuel de la Société canadienne en évaluation. Charlstown.
- Neuendorf, K. A. (2001). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.
- Patton, M. Q. (2001). Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned. *American Journal of Evaluation*, 22(3), 329-336.
- Rescher, N. (1969). *Introduction to Value Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Thousand Oaks: Sage Publications
- Sanders, J. R. (2001). A Vision for Evaluation. *American Journal of Evaluation*, 22(3), 263-266.
- Scriven, M. (1980). *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1995). The Logic of Evaluation and Evaluation Practice. *New Direction for Evaluation*, 68, 49-70.
- Scriven, M. (2001). Evaluation: Future Tense. *American Journal of Evaluation*, 22(3), 301-307.
- Scriven, M. (2006). *Value-driven evaluation*. AEA Conference, Portland, Oregon.
- Secrétariat du Conseil du Trésor du Canada (2004). *Examen de la qualité des évaluations dans les ministères et les organismes*. Document interne, Gouvernement du Canada.
- Shadish, W. E., Cook, T. D., & Leviton, L. C. (1991). *Foundations of Program Evaluation Theories of Practice*. Newbury Park: Sage Publications.
- Silvern, L. C. (1972). *System engineering applied to training*. Houston: Gulf publishing company.
- Stake, R. E. (2004). *Standards-based & Responsive Evaluation*. Thousand Oaks: Sage Publications.
- Toulemonde, J. (2005). *Appropriation des résultats de l'évaluation: Leçons de la pratique en Région Limousin*. Colloque de la Société française de l'évaluation. Lille.
- Toulmin, S. E. (1993). *Les usages de l'argumentation*. Paris: Presses universitaires de France.
- Virtanen, P., & Uusikylä, P. (2002). *Exploring the Missing Links Between Causes and Effects*. Conférence de l'European Evaluation Society, Séville.
- Weiss, C. (1997). *Evaluation: Methods for studying programs and policies*. New Jersey: Pearson Education Canada.