

UN ÉTAT DES LIEUX THÉORIQUES DE L'ÉVALUATION: UNE DISCIPLINE À LA REMORQUE D'UNE RÉVOLUTION SCIENTIFIQUE QUI N'EN FINIT PAS¹

Nathalie Dubois
Richard Marceau
École nationale d'administration publique
Montréal, Québec

Résumé: Cet article effectue l'état des lieux théoriques de l'évaluation de programme en appliquant les principaux critères d'analyse proposés par Kuhn et par Lakatos pour l'étude des disciplines scientifiques. L'étude permet d'identifier les principales composantes théoriques et empiriques de la discipline, de comprendre leurs interactions, et de réfléchir à l'influence qu'elles ont sur le développement de la discipline. L'application du modèle d'analyse conduit au constat que l'évaluation se situe actuellement au troisième stade du développement des disciplines, celui de la révolution scientifique, qui suppose un changement prochain des fondements dominants de la discipline. Cependant, la structuration des connaissances, la relation entre la théorie et la recherche empirique et le schisme épistémologique dominant en sciences sociales concourent à maintenir l'évaluation dans un état de révolution scientifique qui n'en finit pas.

Abstract: This article examines the present state of the theoretical bases of program evaluation by applying key analytical criteria proposed by Kuhn and by Lakatos for the study of scientific disciplines. The study identifies the main theoretical and empirical components of the discipline, clarifies their interactions, and discusses their influence on the development of the discipline. The application of the analytical model leads to the acknowledgement that evaluation is currently at the third stage of development of the disciplines, scientific revolution, which assumes an upcoming change in the discipline's dominant foundations. However, the structuring of knowledge, the relationship between theory and empirical research, and the epistemological schism prevailing in the social sciences contribute to keeping evaluation in a never-ending state of scientific revolution.

Correspondance à l'auteure: Nathalie Dubois, École nationale d'administration publique, 4750 avenue Henri-Julien, 5^e étage, Montréal, Québec H2T 3E5;
<Nathalie_Dubois@enap.ca>

En cours de mandat, les évaluateurs doivent prendre une série de décisions à caractère théorique et méthodologique. Ils peuvent donc s'appuyer sur l'ensemble de leur connaissances théoriques, méthodologiques, et empiriques. Sans cet appui, évaluer un programme demeure un exercice périlleux où la collecte, l'analyse, et l'interprétation des données sont réalisées sans point de repère. La fiabilité et la validité de la démarche évaluative peuvent alors être sérieusement mises en doute. Aussi, l'étude des théories et de la pratique contribue au progrès des connaissances, permet d'améliorer notre pratique, et favorise les débats d'idées au sein d'une discipline.

Cet article étudie l'état actuel des connaissances théoriques et pratiques de l'évaluation des programmes en appliquant un modèle d'analyse du développement des disciplines scientifiques. Ce modèle regroupe plusieurs critères d'analyse proposés par Kuhn (1962/1983) et par Lakatos (1970). Dans son ensemble, l'exercice n'est pas original car plusieurs auteurs ont appliqué ces critères pour mieux comprendre l'évolution ou l'état des connaissances regroupées dans leurs champs d'expertise, par exemple, en sciences économiques, en management, ou encore en sciences politiques. Tout comme pour ces travaux d'analyse, le but premier de cet exercice est d'employer les critères d'analyse comme instruments d'étude afin de contribuer à révéler la structure des connaissances actuelles en évaluation des programmes plutôt que d'apporter une contribution originale à la philosophie des sciences.

Plus concrètement, l'article propose, tout d'abord, une présentation du modèle d'analyse. Ensuite, ce modèle est appliqué à la discipline de l'évaluation de programme. L'exercice permet d'identifier et de schématiser les composantes théoriques et empiriques majeures de la discipline et de mieux comprendre comment elles interagissent. L'application du modèle d'analyse conduit au constat que l'évaluation se situe actuellement au troisième stade du développement des disciplines, celui de la révolution scientifique, qui suppose un changement prochain des fondements dominants de la discipline. Cependant, la structuration des connaissances, la relation entre la théorie et la recherche empirique, et le schisme épistémologique dominant en sciences sociales concourent à maintenir l'évaluation dans un état de révolution scientifique qui n'en finit pas.

UN MODÈLE D'ANALYSE INSPIRÉ DES TRAVAUX DE KUHN ET LAKATOS

Le modèle d'analyse proposé est fortement influencé par les idées de Kuhn (1962/1983) et de Lakatos (1970). En intégrant ces idées dans un seul modèle d'analyse, nous présentons un cadre d'analyse construit autour de quatre préoccupations: (a) les liens unissant la discipline à son environnement (Kuhn), (b) l'évolution de la discipline (Kuhn), (c) l'organisation interne des connaissances théoriques (Lakatos), et (d) la nature des relations unissant les connaissances théoriques aux éléments empiriques de la discipline (Kuhn et Lakatos). L'étude de ces éléments constitue à sa manière un résumé de l'état actuel d'une discipline scientifique et autorise une première évaluation du corpus théorique.

La première considération du modèle concerne les liens unissant la discipline à son environnement. Selon Kuhn (1962/1983), une discipline constitue un système ouvert de connaissances composé de différents éléments ontologiques, épistémologiques, théoriques, et empiriques. Les connaissances scientifiques émanent de la spécificité de chaque discipline (éléments internes) et des disciplines connexes (éléments externes). Ces différents éléments s'influencent mutuellement et la sommation des forces de chacun, à un moment précis, explique l'état de la discipline et la nature plus ou moins multidisciplinaire de celle-ci. Cette structuration des connaissances a un impact direct sur le développement de la discipline et permet d'expliquer l'état actuel des connaissances et la problématique interne et externe de la discipline.

La deuxième considération du modèle se rapporte à l'évolution des disciplines dans le temps. Selon Kuhn (1962/1983), elles traversent quatre phases pendant leur évolution: la pré-science — la science normale — la crise révolutionnaire, et la nouvelle science normale (Chalmers, 1987). À l'aube de leur développement, qui correspond à la première phase de l'évolution, les disciplines possèdent un ensemble désorganisé d'éléments de connaissances. À ce moment, les chercheurs et les praticiens élaborent des principes et des fondements sans réellement s'entendre sur «une vision commune» ou un paradigme spécifique. La fréquence et la rigueur des échanges théoriques et empiriques conduisent au deuxième stade du développement où les connaissances sont davantage organisées autour d'un paradigme dominant accepté par un grand nombre de chercheurs. Ce premier paradigme comprend un ensemble de lois, de modèles

heuristiques, de valeurs, et de prescriptions méthodologiques, ainsi qu'une utilisation établie de certains instruments. Le troisième stade d'évolution correspond à une révolution scientifique. L'intensité des débats tenus au cours du deuxième stade peut conduire les chercheurs et les praticiens à abandonner le premier paradigme pour en adopter un nouveau. Ce changement ne découle pas d'une démonstration claire et rationnelle de l'inefficacité du premier paradigme. Au contraire, Kuhn estime que les chercheurs choisissent d'adhérer à un nouveau paradigme sous l'influence de facteurs psychologiques et sociaux. Cette révolution scientifique est d'une grande importance, car elle traduit un changement profond des principes et des théories. Ce changement conduit les disciplines à un nouvel équilibre, un retour en fait au deuxième stade, où le nouveau paradigme est remis en cause par de nouveaux chercheurs.

La troisième considération du modèle d'analyse concerne l'organisation interne des connaissances théoriques. Selon Lakatos (1970), lorsqu'une discipline atteint le deuxième stade de son développement, elle est constituée de quatre composantes: (a) le noyau dur, (b) l'heuristique négative, (c) l'heuristique positive, et (d) la ceinture de défense. Ces éléments décrivent le programme de recherche qui équivaut au concept de «paradigme» défini préalablement par Kuhn. Plus précisément, le noyau dur correspond aux fondements ontologiques, épistémologiques, et théoriques. Le noyau dur est mis à l'abri des tests empiriques par un ensemble de prohibitions méthodologiques (heuristique négative). Ces hypothèses sont regroupées dans la ceinture de défense, qui entoure et protège le noyau dur, et sont soumises aux essais empiriques. La ceinture de défense se développe grâce à l'heuristique positive, c'est-à-dire un ensemble de théories et de méthodologies constitué dans le but d'appuyer les fondements du noyau dur. Selon Lakatos (1970), l'heuristique négative indique que le noyau dur est réfutable par la discussion méthodologique des protagonistes alors que l'heuristique positive comprend l'ensemble des suggestions de changements pouvant être apportés à la lumière des derniers résultats empiriques.

La quatrième considération se rapporte à la nature des relations unissant les connaissances théoriques et empiriques de la discipline. Selon les deux auteurs, les connaissances théoriques doivent être mises en interaction avec le monde réel de manière à confronter le paradigme ou le programme de recherche à la réalité. Le terrain empirique constitue la réalité sujette à l'observation et à l'explication. La relation entre la théorie et la recherche empirique corres-

pond, premièrement, à l'application des heuristiques négatives à la pratique et, deuxièmement, à la formulation de nouvelles heuristiques positives influencées par les résultats de la recherche. Cette relation correspond à un développement progressif des connaissances. Sans cette relation, le programme de recherche (ou paradigme) demeure stable, sans changement, sans innovation significative pour la discipline. Dans cette situation, le développement théorique est considéré comme dégénératif, c'est-à-dire qu'aucune confirmation ou infirmation des propositions n'est produite à la lumière des travaux empiriques. De cette façon, les deux niveaux de connaissances, se développent sans liens évidents.

En définitive, le modèle d'analyse révèle une dynamique où les éléments de connaissances naissent, se développent, s'influencent, ou quelquefois sont rejetés. Cette dynamique suggère deux niveaux de connaissances: (a) le niveau des fondements théoriques, ainsi que (b) le niveau empirique. Le premier niveau de connaissances correspond au noyau dur (heuristique négative) et à la ceinture de défense (heuristique positive) alors que le second niveau désigne le terrain empirique (Figure 1). Ces deux niveaux de connaissances doivent être en interaction continue et évolutive, ce qui donne lieu à un développement conjoint des deux niveaux de connaissance. Par conséquent, le niveau empirique est sujet à l'influence des percées théoriques et ce mouvement de changements est aussi observable au premier niveau des connaissances au moment où la pratique produit de nouveaux résultats empiriques. Cette évolution des connaissances est dite progressive et souhaitable.

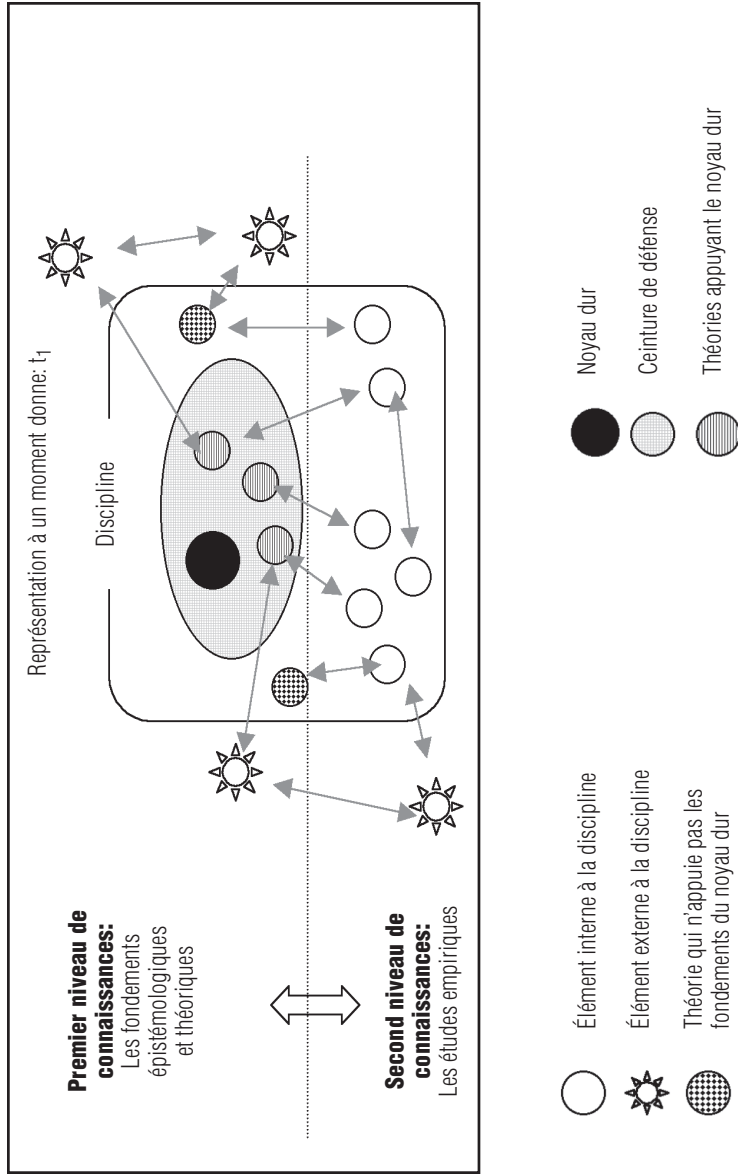
APPLICATION DU MODÈLE D'ANALYSE À L'ÉVALUATION DE PROGRAMME

La présente section applique le modèle d'analyse à l'évaluation de programme. La discipline actuelle est étudiée en regard des quatre considérations du modèle d'analyse. Cette application du modèle constitue la base de l'analyse critique présentée par la suite.

Selon la première considération du modèle, une discipline constitue un système ouvert regroupant des connaissances internes et externes à la discipline. Il est aisé d'identifier deux types de connaissances en évaluation de programme.

Tout d'abord, plusieurs éléments internes émanent directement de la discipline. Ainsi, l'évaluation de programme possède son propre

Figure 1
Une conception du développement des disciplines scientifiques à la deuxième phase (stabilité de la discipline)



langage constitué d'un riche vocabulaire (méta-évaluation, *empowerment evaluation*, *value-free evaluation*, évaluation formative, évaluation sommative, etc.) qui alimente les théories, les concepts, et les aspects techniques de la discipline. Le vocabulaire évaluatif a été répertorié et documenté entre autres par Scriven (1991). Depuis *Evaluation Thesaurus* n'est plus la seule référence sur le sujet, mais elle montre de manière éloquente la richesse des connaissances internes de cette discipline.

Des éléments externes influencent aussi la discipline. Ces connaissances proviennent entre autres du domaine de l'éducation, de la sociologie, de la psychologie, et de l'économie. Théoriquement, ces éléments externes influencent continuellement les connaissances internes permettant ainsi à la discipline d'évoluer. En évaluation de programme, cette influence demeure toutefois sectorielle, ce qui limite l'ampleur de son influence sur l'évolution de la discipline. Par exemple, les psychologues et les éducateurs qui pratiquent l'évaluation gravitent autour de l'*American Evaluation Association*, les économistes et les politicologues se réunissent à l'*American Association of Public Policy and Management*, et les chercheurs en santé regroupent leurs travaux dans différentes associations relatives à l'éducation sanitaire, l'épidémiologie, la biostatistique, ou la santé publique (Shadish, Cook, & Leviton, 1991). Cette fragmentation des connaissances s'observe aussi au niveau des publications. Bien que plusieurs revues se consacrent à l'évaluation de programme (e.g., *Evaluation Review*; *Evaluation Practice*; *Evaluation: International Journal of Theory, Research and Practice*; *Evaluation and Program Planning*; *New Directions for Evaluation*; *Canadian Journal of Program Evaluation*; *Evaluation Journal of Australasia*), des articles significatifs pour la discipline sont souvent publiés par des revues issues de disciplines connexes à l'évaluation (e.g., *Evaluation and Health Professions*, *Educational Evaluation and Policy Analysis*, *Journal of Policy Analysis and Management*, etc.). Cette situation conduit à un éclatement des connaissances, certes, mais aussi à un éparpillement des progrès réalisés.

L'étude de la première considération du modèle d'analyse révèle également la perméabilité de l'évaluation aux connaissances issues d'autres disciplines. Cette ouverture de la discipline à son environnement nourrit très clairement les concepts, les théories, et les méthodologies retenues en évaluation de programme. Il demeure cependant difficile de juger si la discipline est multidisciplinaire ou si elle représente un amalgame de plusieurs disciplines. Il importe

pourtant de s'entendre sur une définition de l'évaluation. Une formulation claire des caractères et une détermination exhaustive du contenu relié à l'évaluation de programme pourrait contribuer à réduire l'isolement actuel des débats théoriques et empiriques dans les disciplines d'origine et ainsi promouvoir le développement d'une discipline plus cohérente et intégrée.

La deuxième considération du modèle concerne l'évolution de la discipline. Selon le modèle d'analyse, les disciplines traversent différents stades durant leur développement. Au gré des changements structuraux des éléments internes et externes, la discipline traverse trois stades décrivant, à un moment précis dans le temps, l'état des connaissances. L'objet de la présente section est d'identifier le stade actuel du développement de l'évaluation. Ce portrait permet d'identifier dans quelle phase du développement l'évaluation de programme se situe et, ainsi, de mieux comprendre les problématiques académiques et professionnelles qui en découlent.

Selon Madaus, Scriven, et Stufflebeam (1983), l'évaluation se pratique depuis le 19^e siècle. Depuis ce temps, nous estimons que l'évaluation de programme a franchi les trois stades de développement théorique et empirique. Ainsi, à l'aube de son développement, la discipline possédait un ensemble désorganisé d'éléments de connaissances (phase 1). À ce stade, l'évaluation de programme était une discipline désorganisée où différents auteurs publiaient des ouvrages fournissant les premières bases de la discipline. Ce désordre dans les éléments théoriques et empiriques est néanmoins jugé constructif, car il a fourni certains concepts élaborés ultérieurement dans le paradigme post-positiviste. Parmi ces concepts, on peut penser à la mesure objective et à la détermination d'objectifs explicites.

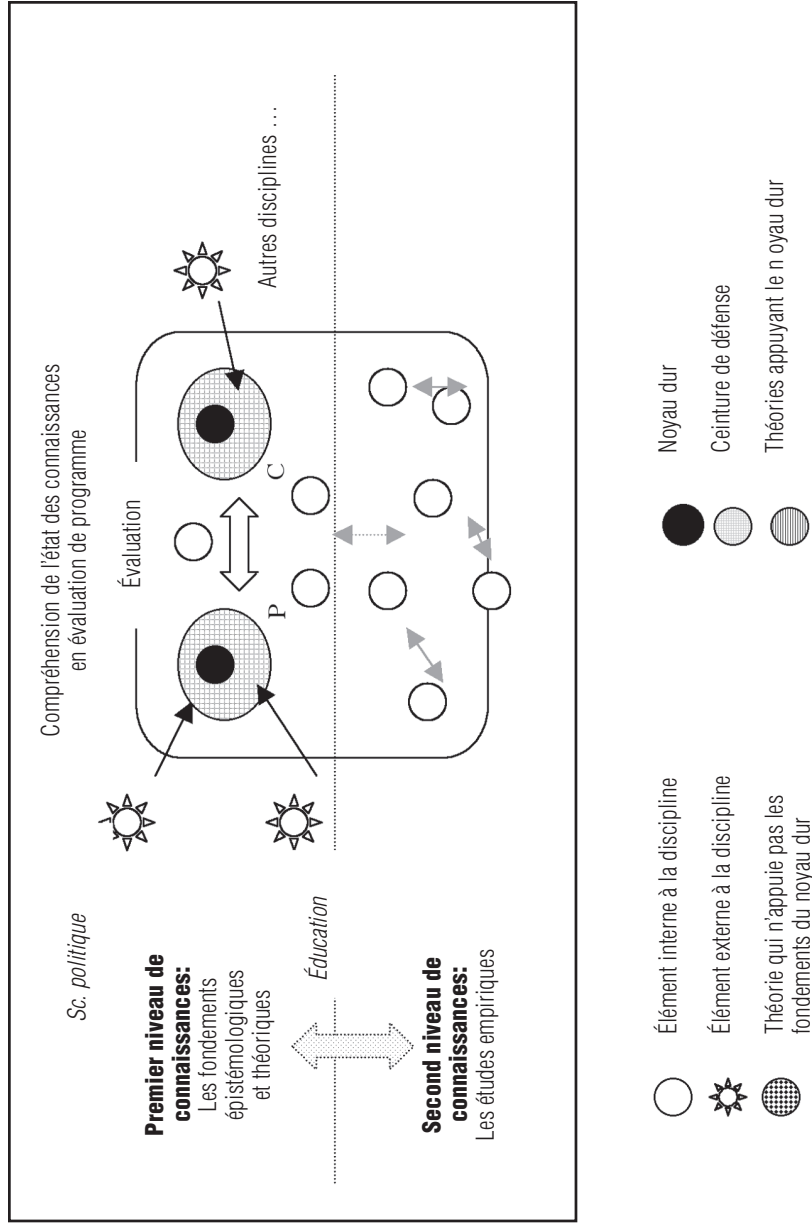
Lors de sa «professionnalisation», la discipline a atteint la seconde étape de développement puisqu'elle est devenue plus stable en raison de l'organisation croissante des connaissances autour des fondements post-positivistes. Durant cette phase, l'évaluation de programme a connu une période de stabilité à l'abri des discordes internes. La discipline était alors constituée de trois composantes (noyau dur, ceinture de défense, et terrain empirique) qui se regroupaient en deux niveaux de connaissances. Les éléments qui constituaient alors ces deux niveaux de connaissances étaient en interaction. Cette relation constante a permis un développement conjoint des deux niveaux de connaissance et a contribué fort probablement à conduire l'évaluation de programme vers le troisième

stade de développement des disciplines, celui de la révolution scientifique. À ce stade, des idées constructivistes ont vu le jour en sciences sociales et ont également influencé profondément le débat en évaluation. Plusieurs auteurs ont participé à la construction de ce que l'on pourrait appeler un noyau dur constructiviste et une ceinture de défense. Ces nouveaux éléments de connaissances obéissent à des fondements nettement distincts du premier noyau dur mais, comme pour le paradigme post-positiviste, les chercheurs et praticiens constructivistes élaborent une argumentation présentant leur perception du monde et des connaissances que l'évaluation de programme devrait privilégier (Figure 2). Au cœur de ce paradigme repose en fait une tout autre idée de la science.

Selon Kuhn (1962/1983), à mesure que davantage de savants se convertissent, pour diverses raisons d'ailleurs, au nouveau paradigme, il se produit une modification croissante de la distribution des persuasions professionnelles. Si la révolution est victorieuse, cette modification va faire tache d'huile jusqu'à concerner la majorité de la communauté scientifique, ne laissant à l'écart qu'une poignée de dissidents (Chalmers, 1987). Théoriquement, la richesse des débats tenus au cours du deuxième stade devrait en principe conduire les universitaires et les professionnels à abandonner le premier noyau dur pour adopter le nouveau qui correspond davantage à la discipline. Ce changement permet à la discipline d'atteindre un nouvel équilibre, un retour au deuxième stade de développement.

L'analyse des articles, des rapports de recherche et des débats en évaluation laisse présumer que l'évaluation de programme se situe toujours au troisième stade de développement (période de révolution scientifique). Depuis l'émergence du second paradigme, aucun débat n'a permis au noyau dur constructiviste de déloger le noyau dur post-positiviste ou de limiter l'influence du premier noyau sur le corpus des connaissances. Une analyse plus précise des deux niveaux de connaissances (théorique et empirique) vise à mieux comprendre les raisons pour lesquelles, après autant d'années, l'évaluation de programme possède toujours deux noyaux durs, deux programmes de recherche, au sein de son corpus théorique. Cette étude inclut l'analyse des deux dernières considérations du modèle, c'est-à-dire l'organisation interne des connaissances théoriques et la nature des relations unissant les connaissances théoriques et empiriques de la discipline.

Figure 2
Compréhension de l'état des connaissances en évaluation de programme



Le développement du niveau théorique de connaissances en évaluation de programme

Le premier niveau de connaissances renvoie aux fondements théoriques regroupés dans le noyau dur et la ceinture de défense de la discipline. Plus précisément, le noyau dur est constitué des éléments fondamentaux de la discipline (ontologique et épistémologique) et peut se comparer à une masse dense où l'évolution des éléments qui le constituent est lente et marginale. La ceinture de défense représente, quant à elle, un milieu plus dynamique où les auteurs proposent et défendent différentes théories et hypothèses qui appuient et solidifient le noyau dur (heuristiques positives). Par conséquent, elle constitue un milieu perméable entre le noyau dur et le terrain empirique ayant comme principal objectif de défendre les fondements du noyau dur (heuristiques négatives) tout en les confrontant aux réalités pratiques de la discipline.

L'étude du premier niveau des connaissances nécessite une analyse des deux noyaux durs constituant actuellement la discipline (évaluation post-positiviste et évaluation constructiviste). Pour chacun d'eux, nous proposons, premièrement, d'étudier les fondements du noyau dur ainsi que les principales variances théoriques acceptées par la communauté scientifique. Ces variances visent à adapter les fondements proposés par le noyau à la pratique en évaluation sans pour autant les trahir. Deuxièmement, l'article identifie les principales théories, modèles, ou approches pouvant loger dans chacune des ceintures de défense (post-positiviste et constructiviste). La présentation de ces théories sera brève car l'objectif n'est pas de dresser un portrait détaillé de chacune d'entre-elles (certains auteurs ont déjà accompli ce travail dont, entre autres, Shadish, Cook, & Leviton, 1991; Stufflebeam, 2001) mais plutôt de mieux comprendre la structure actuelle des connaissances et son influence sur le développement de la discipline.

Les schématisations des principales connaissances théoriques regroupées dans les ceintures de défense (Figures 3 et 4) correspondent à une réorganisation des modèles théoriques définis par Stufflebeam (2001). Des 22 approches proposées par l'auteur, seules les théories associées à l'évaluation de politiques ou de programmes publics sont considérées. Ainsi, cinq approches sont exclues de la présente analyse, car elles ne permettent pas d'analyser les politiques et les programmes publics comme résultante de l'activité publique.

Un premier noyau dur et sa ceinture de défense: L'évaluation post-positiviste

Le premier noyau dur identifiable en évaluation est de nature post-positiviste et est constitué de trois principaux fondements. Le premier fondement est que l'évaluation possède sa propre réalité. Suivant les enseignements de Mills (1967), les auteurs considèrent que l'évaluation existe par la connaissance des faits et de l'expérience scientifique (Scriven, 1991). On propose alors une approche déductive où les théories guident la recherche empirique et les résultats empiriques nourrissent le développement des connaissances théoriques. L'évaluation est ainsi réalisée à la base de modèles qui orientent l'évaluateur dans son choix de variables dépendantes et indépendantes. Cette façon de procéder permet à l'évaluateur de présenter la réalité de l'objet de façon acontextuelle et objective (Campbell & Stanley, 1966; Cook & Campbell, 1979; Scriven, 1980; Stake, 1975; Suchman, 1967).

Bien que fidèles au noyau dur post-positiviste, certains auteurs considèrent important d'adapter les exigences scientifiques de la recherche post-positiviste aux réalités de la pratique en évaluation (Weiss, 1997). Les auteurs souhaitent donc que l'évaluation réponde aux exigences de la recherche scientifique tout en étant conscients qu'elle est dans certains cas un acte politique qui se manifeste dans un contexte où le pouvoir, les idéologies, et les intérêts sont souvent plus puissants que la prise de décision rationnelle (Rossi, Freeman, & Lipsey, 1999; Weiss, 1977; Weiss & Buccuvalas, 1980; Wholey, 1979). Ainsi, la démarche évaluative doit s'ajuster à la réalité de la pratique sinon elle demeura oiseuse pour le décideur public (Weiss, 1977).

Un deuxième fondement du noyau dur positiviste est l'amélioration du bien-être de la collectivité par l'intermédiaire de l'évaluation des interventions étatiques. Le but de l'évaluation est de mesurer l'efficacité et l'efficience de l'action publique afin d'améliorer le bien-être collectif des membres de la société. Par conséquent, les données recueillies sont planifiées et organisées pour répondre à cette conception de l'évaluation (Cronbach, 1963). Stake (1967) précise que l'évaluation doit décrire les relations qui existent entre l'intervention publique et sa performance sociale. De plus, Cronbach (1980) ajoute que l'évaluation doit guider le développement d'un programme et permettre ainsi une prise de décision plus fréquente de la part des citoyens. Pour cela, l'évaluation doit permettre de porter un jugement global sur le programme ou la politique. Ce dernier est établi sur la base de mesures,

de statistiques, ou de stratégies expérimentales qui s'appuient sur une bonne compréhension du programme (Scriven, 1991). Ce jugement implique une estimation des différences entre le pré- et le post-test, entre les groupes expérimentaux et les groupes contrôles, ainsi que sur un nombre de paramètres critiques.

À ces propositions, certains auteurs suggèrent que l'évaluation puisse permettre d'analyser les activités entreprises afin d'apprécier les modalités et les effets des politiques publiques et des programmes (Bernstein & Freeman, 1975; Chen & Rossi, 1980). Cet approfondissement permet à l'évaluateur d'expliquer comment le programme fonctionne, pourquoi il fonctionne de cette façon, et comment l'efficacité de ce dernier peut être améliorée.

Un troisième fondement de l'évaluation post-positiviste est l'indépendance entre l'objet et l'évaluateur. Cette indépendance découle de la conception objective et acontextuelle de la connaissance. Le concept d'objectivité permet, selon Popper (1991), de supposer que l'observateur ne peut modifier la nature de l'objet étudié et, ainsi, ne peut être influencé par les opinions personnelles, les goûts, et les spéculations de l'imagination. Cette indépendance permet à l'évaluateur d'accorder une valeur objective à l'objet en utilisant des standards de mérite (Scriven, 1980). Cette conception de la valeur conduit Scriven (1967) à développer la dimension du jugement en évaluation. Il définit ce jugement comme étant «une activité méthodologique qui consiste à combiner les données ou les résultats d'une action avec une échelle d'objectifs» (Scriven, 1980). Ainsi, toute évaluation implique un jugement de valeur en guise de conclusion et ce rôle de juge est octroyé à l'évaluateur pour divers motifs. Le premier motif, selon Scriven (1991), Suchman (1967), et Campbell et Stanley (1966), est que l'évaluateur possède une formation spécialisée en évaluation. Le deuxième motif est que l'évaluateur est un acteur externe à l'environnement de l'activité à évaluer; il n'a donc aucun jugement préétabli en fonction des valeurs ou des connaissances personnelles (Scriven, 1980). Le troisième motif concerne les méthodes et les techniques de recherche utilisées qui peuvent assurer à l'évaluateur une certaine rigueur scientifique (Campbell & Stanley, 1966; Cook & Campbell, 1979; Scriven, 1980). Enfin, l'évaluateur porte un jugement sur la valeur des programmes à la lumière d'études et d'analyses rigoureuses. Cette conception du rôle de l'évaluateur permet l'établissement de la notion de «professionnel en évaluation» qui suppose que l'évaluateur est le seul acteur pouvant produire une évaluation juste et faire des recommandations aux demandeurs de l'évaluation.

D'autres réflexions sont proposées quant à l'application de l'indépendance de l'évaluateur face à l'objet, mais celles-ci sont toutefois de nature plus souple et conduisent à un changement du rôle de l'évaluateur (Chen & Rossi, 1989; Rossi, Freeman, & Lipsey, 1999). Ce changement suggère la participation d'un plus grand nombre d'acteurs aux processus évaluatifs. Selon cette logique, l'évaluateur qui intègre les préoccupations, les intérêts, et les valeurs des personnes touchées par le programme bonifie son évaluation. Tout comme Cronbach (1963, 1980), Rossi et Freeman (1985) encouragent les professionnels à tenir compte des valeurs et des considérations des gestionnaires et décideurs, mais comme Scriven (1991) et Campbell et Stanley (1966), ils croient que cette considération ne doit pas entraver la rigueur de la démarche d'évaluation ainsi que le jugement du professionnel quant à l'élément évalué. Cette nouvelle conception du rôle de l'évaluateur laisse présumer que les conclusions de l'évaluation peuvent être un peu moins objectives et quelque peu négociées.

Tableau 1
Résumé des principaux fondements du noyau dur post-positiviste et des théories dominantes élaborées par la ceinture de défense

Les fondements du noyau dur post-positiviste	Les théories dominantes développées par la ceinture de défense
1. L'évaluation possède une essence propre à elle-même	<ul style="list-style-type: none"> • Constituée de faits et d'expériences empiriques
2. La finalité de l'évaluation: expliquer	<ul style="list-style-type: none"> • Évaluation pour les décideurs publics de haut niveau et pour les membres de la communauté (les citoyens) • Recherche de l'explication par la théorie et l'empirisme • Mesure de l'efficacité et de l'efficience de l'intervention • Jugement et proposition de solutions en vue d'une amélioration
3. L'évaluateur est indépendant de l'objet d'évaluation, car la connaissance est objective et acontextuelle	<ul style="list-style-type: none"> • Les questions d'évaluation: Est-ce que le programme a procuré les résultats attendus? Est-ce que les résultats sont dus au programme? Est-ce que le programme représente la solution la plus efficace pour régler le problème? • Protocole d'évaluation structuré (théories, hypothèses, variables dépendantes et indépendantes, etc.) • Vérification des hypothèses et généralisation des résultats • Devis expérimental ou quasi expérimental avec données avant et après intervention et groupe(s) contrôle(s) • Méthodes quantitatives • Données principalement quantitatives mais peuvent être, au besoin, des données qualitatives

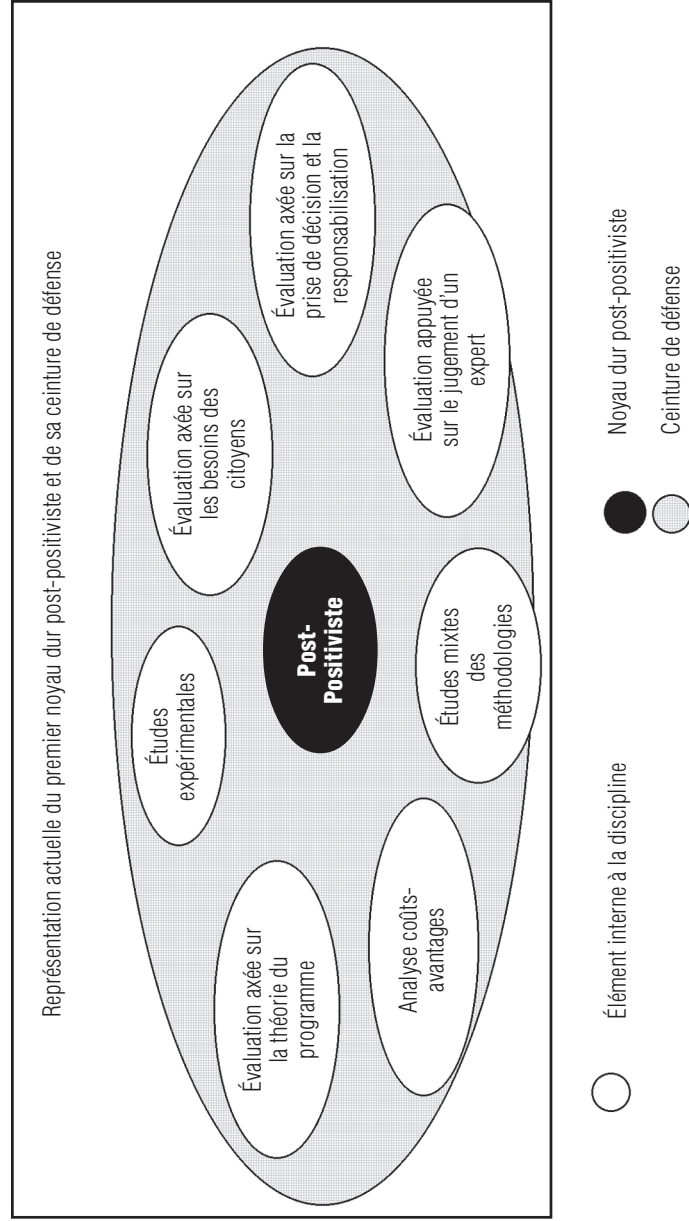
De ce résumé, il est possible d'identifier, parmi les théories proposées par Stufflebeam (2001), six approches qui défendent, en totalité ou en grande partie, les fondements du noyau dur post-positiviste (Figure 3). Ces théories sont associées à deux groupes d'approches visant à mesurer le mérite et la valeur d'un programme: (a) les approches évaluatives orientées vers une amélioration des programmes et une responsabilisation plus grande de la part des décideurs publics et (b) les approches évaluatives guidées par des questions ou des méthodologies.

Plus précisément, au cœur du premier groupe d'approches, on retrouve, premièrement, une théorie qui se concentre sur les besoins des citoyens (*consumer-oriented studies*). Selon cette théorie, l'évaluation est un processus qui permet de déterminer le mérite et la valeur du programme où le bien-être du citoyen demeure la justification première du programme. Selon les propositions de Scriven (1967, 1991, 1994), Glass (1975), et Cronbach (1963), l'évaluation permet de juger les bénéfices retirés et les coûts assumés par les citoyens (Stufflebeam, 2001). Ce jugement est établi sur la base de mesures, de statistiques, ou de stratégies expérimentales qui s'appuient sur une bonne compréhension du programme (Scriven, 1996). Ce jugement prend toutefois des formes diverses (e.g., l'évaluation formative et l'évaluation sommative proposées par Scriven, 1991). De plus, cette théorie propose la méta-évaluation qui correspond à l'évaluation d'une évaluation (Scriven, 1980). L'intérêt premier de la méta-évaluation est de contrôler le plus grand nombre de biais possibles lors de l'évaluation et, par conséquent, de favoriser un plus grand bien collectif. La deuxième théorie émergeant de ce premier groupe concerne l'évaluation orientée vers la prise de décision et la responsabilisation des actions publiques (*decision/accountability-oriented evaluation approach*). Selon les défenseurs de cette théorie, l'évaluation doit être employée proactivement de manière à améliorer continuellement le programme et à juger rétroactivement son mérite et sa valeur. Pour les pionniers de cette approche (Cronbach, 1963; Stufflebeam, 1966, 1967), l'évaluation vise à informer le décideur public quant aux services rendus aux citoyens et à le rendre responsable des répercussions de ses décisions.

Le second groupe propose cinq théories permettant de mesurer le mérite et la valeur d'un programme à l'aide de questions ou de méthodologies précises. La première théorie concerne les «études expérimentales» (*experimental studies*) qui déterminent les rapports causaux entre des variables indépendantes et dépendantes afin de



Figure 3
Une conception du premier noyau dur et sa ceinture de défense post-positiviste (selon les modèles théoriques en évaluation proposés par Stufflebeam, 2001)



répondre à des questions précises d'évaluation. Conscients des réalités auxquelles l'évaluation est confrontée, les auteurs proposent d'utiliser les devis quasi-expérimentaux et expérimentaux comme méthodologie d'évaluation afin de limiter les biais associés à la recherche (Campbell & Stanley, 1963; Cronbach & Snow, 1969; Lindquist, 1953; Suchman, 1967). La deuxième théorie suggère de réaliser l'évaluation des programmes à partir d'une conception théorique du programme (*program theory-based evaluation*). Cette approche propose de schématiser les liens unissant le programme aux autres variables de la problématique d'intervention afin de mieux comprendre les raisons du succès ou de l'échec du programme. Les pionniers de cette approche sont Glaser et Strauss (1967), Weiss (1972, 1995), Bickman (1990), Chen (1990), et Rogers (2000). La troisième théorie concerne l'analyse coûts-avantages (*cost-benefit analysis approach*) qui permet de mesurer les ressources investies, de comprendre comment elles ont été investies, et d'estimer leurs effets. Cette analyse économique permet de déterminer et de juger si les investissements ont permis d'atteindre efficacement les objectifs sociaux convenus. Selon Stufflebeam (2001), les principaux auteurs qui ont participé à l'édification de cette théorie sont Kee (1995), Levin (1983), et Tsang (1997). La quatrième théorie s'appuie sur le jugement critique d'un expert provenant d'un domaine particulier (*criticism and connoisseurship*). Cette théorie, inspirée des méthodes employées dans la critique d'art et littéraire, suppose que certains experts peuvent, suite à une analyse détaillée, attribuer une valeur au programme. Ce jugement critique s'appuie sur la sensibilité, l'expérience, et les connaissances précises de l'expert (Eisner, 1975, 1983; Flinders & Eisner, 2000; Vallance, 1973). La cinquième théorie (*mixed-methods studies*) propose de recourir à plusieurs méthodes (qualitatives et quantitatives) afin d'évaluer l'efficacité d'un programme. Selon les auteurs qui ont élaboré cette théorie (Guba & Lincoln, 1981; Kidder & Fine, 1987; Lincoln & Guba, 1985; Miron, 1998; Patton, 1990; Schatzman & Strauss, 1973; Tyler, 1932), l'évaluateur doit tirer profit de différents outils qui permettent d'obtenir des informations et ainsi d'évaluer plus précisément le mérite et la valeur d'un programme.

Un second noyau dur: L'évaluation constructiviste

Inspirées du mouvement critique des connaissances, de nouvelles idées voient le jour et influencent les débats en évaluation. Ces propositions, réunies dans le noyau dur constructiviste, répondent à trois principaux fondements ontologiques et épistémologiques. Tout

d'abord, l'évaluation ne pourrait rapporter objectivement la réalité. Selon Guba et Lincoln (1989), la réalité décrite par les post-positivistes n'existe pas. Pour ces auteurs, il y a autant de réalités qu'il y a d'individus, car la réalité correspond à une construction mentale spécifique.² L'évaluation ainsi que ses conclusions ne doivent pas représenter uniquement le jugement objectif de l'évaluateur, mais elles doivent plutôt inclure et correspondre aux valeurs dominantes des acteurs engagés dans l'évaluation (Weiss, 1972). Par sa participation directe dans le programme et l'environnement du programme, l'évaluateur est en mesure de négocier et de construire un sens commun relatif au programme (Guba & Lincoln, 1989). Par conséquent, le chemin de la connaissance emprunte nécessairement celui de la méthode inductive.

Un deuxième fondement qui caractérise la réflexion des constructivistes est la réalité dite «expérimentable». Selon cette logique, l'évaluateur est en médiation constante avec les valeurs, les attitudes, les croyances, et les significations que chaque personne accorde à l'objet étudié. Pour cette raison, l'évaluation présente davantage les résultats d'une négociation entre différents intervenants qu'un jugement objectif de l'évaluateur (Guba & Lincoln, 1989). L'évaluateur devient donc un partenaire dans la réalisation de l'évaluation: il guide les personnes intéressées par l'évaluation et cela principalement sur le plan des méthodes et des techniques. De plus, le jugement évaluatif qui est porté à la fin de toute évaluation est réalisé autant par l'évaluateur que par l'ensemble des autres acteurs engagés dans l'évaluation (Guba & Lincoln, 1989).

Un troisième fondement qui caractérise la réflexion constructiviste est le désir de développer une compréhension du programme selon l'environnement et le contexte. Les personnes intéressées par cette compréhension du programme (les demandeurs de l'évaluation) sont les responsables du programme, les employés, ainsi que les bénéficiaires (Greene, 2000; Guba & Lincoln, 1989; Patton, 1997a, 1997b). Pour eux, une évaluation de programme doit permettre de comprendre le programme et d'améliorer continuellement ses faiblesses. Pour cela, l'évaluateur élabore une méthodologie réalisée directement dans les milieux naturels, car ces milieux regroupent les différentes forces politiques et sociales qui influencent l'objet étudié et qui ne peuvent pas être reproductibles dans un environnement stérile. Par conséquent, les méthodologies préconisées par les évaluateurs constructivistes sont de formes flexibles, peu organisées, et évolutives. Au fur et à mesure que l'évaluation progresse, l'évaluateur construit sa méthodologie et celle-ci dépend de ses perceptions et de ses analyses.

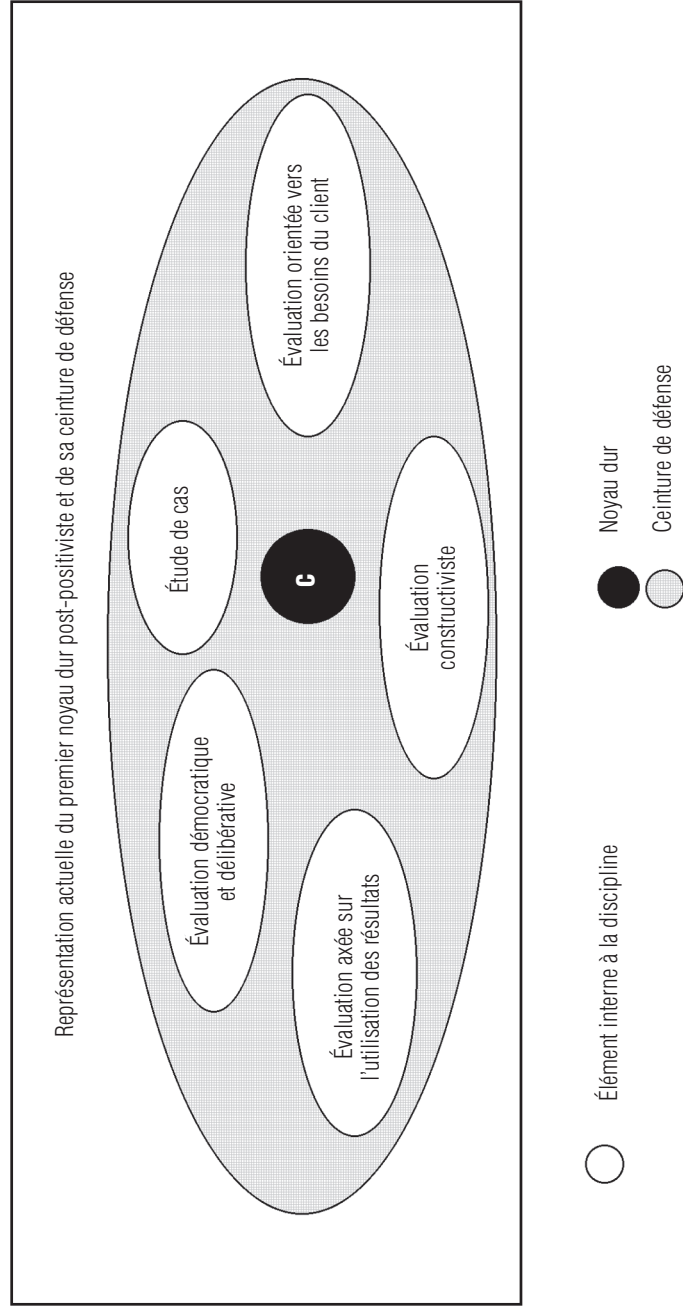
Néanmoins, l'étude du programme et de son environnement doit répondre à certaines règles afin d'en assurer la qualité: la crédibilité, la transférabilité, la fiabilité, et la dépendance (Guba & Lincoln, 1982). La crédibilité s'intéresse à la représentativité des sources de données et elle vérifie si la formulation de l'analyse et de l'interprétation est crédible du point de vue des acteurs auprès desquels on effectue les validations des constructions proposées par l'évaluateur. La transférabilité veille à la bonne description des circonstances et à la capacité de poser un jugement raisonnable quant au degré de transférabilité. La fiabilité précise les décisions de l'évaluateur sur les adéquations en cours de route du processus d'évaluation. La dépendance vise à s'assurer que la construction de l'évaluateur n'a pas de prédominance sur la construction des autres acteurs.

Il est possible d'identifier parmi les théories proposées par Stufflebeam (2001) cinq théories qui défendent, en totalité ou en partie, les fondements du noyau dur constructiviste (Figure 4). Quatre théories accordent une place de choix aux différents acteurs associés au programme alors qu'une seule théorie permet d'étudier le programme à l'aide d'une question ou d'une méthodologie précise.

Tableau 2
Résumé des principaux fondements du noyau dur constructiviste et des théories dominantes développées par la ceinture de défense

Les fondements du noyau dur constructiviste	Les théories dominantes développées par la ceinture de défense
1. L'évaluation est une réalité pluraliste	<ul style="list-style-type: none"> • Raisonnement inductif • Expérience personnelle et construite
2. La finalité de l'évaluation: comprendre	<ul style="list-style-type: none"> • Évaluation pour les responsables du programme, les employés, et les bénéficiaires du programme • Compréhension du programme et de son environnement • Jugement collectif et proposition de solutions en vue d'une amélioration
3. L'évaluateur est dépendant de l'objet d'évaluation, car la connaissance est subjective et contextuelle	<ul style="list-style-type: none"> • Les questions d'évaluation: Comment le programme est-il perçu? Quelles sont les forces et les faiblesses du programme? • Protocole d'évaluation non structuré • Extrapolation des résultats • Stratégie d'évaluation souple, flexible, et évolutive • Instrument: l'humain! • Méthodes qualitatives • Données principalement qualitatives

Figure 4
Une conception du second noyau dur et de sa ceinture de défense constructiviste (selon les modèles théoriques en évaluation proposés par Stufflebeam, 2001)



Plus précisément, le premier groupe d'approches propose quatre théories évaluatives accordant, dans la démarche évaluative, une place de choix aux différents acteurs associés au programme, favorisant une orientation de recherche constructiviste et l'utilisation de méthodes qualitatives (Stufflebeam, 2001). La première théorie concernée correspond aux études orientées vers le client du programme (*client-centered studies, responsive evaluation*). Selon cette théorie, l'évaluation est réalisée en partenariat avec les acteurs qui soutiennent, développent, et administrent le programme. Contrairement aux théories post-positivistes qui se concentrent sur les besoins des citoyens (*consumer-oriented studies*), cette théorie n'a aucune attente quant aux concepts d'indépendance et d'objectivité. La deuxième théorie concerne l'évaluation constructiviste proprement dite (*constructivist evaluation*). Celle-ci considère que l'évaluation doit engager différents acteurs dans le processus d'évaluation et qu'elle ne doit pas demeurer l'apanage des professionnels en évaluation. La troisième théorie concerne l'évaluation démocratique et délibérative (*deliberative democratic evaluation*) proposée par House et Howe (1998). Cette jeune théorie propose d'utiliser le cadre démocratique pour consolider les conclusions d'une évaluation d'un programme. La quatrième théorie, découlant des travaux de Patton (1997a, 1997b, 2000), propose une évaluation qui assure l'utilisation des résultats (*utilization-focused evaluation*). Tous les aspects de l'évaluation sont sélectionnés et appliqués de manière à aider les utilisateurs à obtenir de l'information relativement au programme et à utiliser à court terme les résultats de l'évaluation.

Enfin, la cinquième théorie concerne l'évaluation par étude de cas (*case study evaluations*) et correspond à une approche évaluative permettant d'étudier le programme à l'aide d'une question ou d'une méthodologie précise. L'étude de cas fournit aux acteurs touchés par le programme (*stakeholders*) une interprétation fondée, détaillée et bien documentée du programme. Cette étude ne permet toutefois pas de juger du mérite et de la valeur sociale du programme. Selon Stufflebeam (2001), les principaux pionniers de cette théorie sont Campbell (1975), Lincoln et Guba (1985), Platt (1992), Smith et Pohland (1974), Stake, Easley, et Anastasiou (1978), Stake (1995), ainsi que Yin (1992).

Des théories qui gravitent autour des noyaux durs post-positiviste et constructiviste

Notre analyse de la discipline a permis de constater que, parmi les théories proposées par Stufflebeam (2001), trois théories se retrou-

vent à l'extérieur des deux ceintures de défense parce que leurs hypothèses centrales ne sont pas conformes aux fondements des noyaux post-positiviste ou constructiviste. Bien que ces théories soient guidées par des questions ou des méthodologies permettant d'étudier un aspect précis du programme, nous pensons qu'elles s'apparentent beaucoup plus à «l'analyse managériale des programmes» qu'à l'évaluation.

Selon Stufflebeam (2001), la première théorie proposée (*objectives-based studies*) développée principalement par les travaux de Tyler s'intéresse particulièrement à l'étude des objectifs du programme. Selon cette théorie, l'évaluateur doit formuler et évaluer l'atteinte des objectifs fixés par le client et spécifiés par le gestionnaire du programme. Bien qu'intéressante, l'évaluation se concentre principalement sur les extrants du programme plutôt que sur les effets. La deuxième théorie vise à accroître la responsabilité du gouvernement envers la mise en œuvre de ses programmes et du rendement de ces derniers (*accountability, particularly payment by results studies*). Pour cela, l'évaluateur étudie les effets du programme d'une perspective managériale et organisationnelle. En fonction des résultats managériaux, l'évaluateur propose des sanctions pour ceux qui ont obtenu de faibles résultats et des récompenses pour ceux qui présentent de bons résultats. La troisième théorie propose un système intégré de gestion (*management information systems*) fournissant les informations nécessaires à la gestion et la reddition de compte des programmes. Cette théorie s'apparente à la théorie post-positiviste (*decision/accountability-oriented*) mais fournit uniquement des informations à la direction lui permettant d'appliquer et de rendre compte de sa gestion des programmes.

Le second niveau de connaissances: La recherche empirique

En permettant des échanges continus entre les fondements du noyau dur défendus par la ceinture de défense et les réalités vécues par les praticiens, les études empiriques permettent à la discipline d'évoluer et de s'adapter aux nouvelles préoccupations de l'évaluation de programme. Au Canada, il existe deux façons d'évaluer les programmes qui se distinguent principalement quant à l'utilité première des travaux d'évaluation: (a) une évaluation utile à l'avancement des connaissances disciplinaires et (b) une évaluation utile à la prise de décision. L'étude de Shadish et Epstein (1987) confirme la présence de ces deux conceptions en évaluation. Plus précisément, ceux-ci constatent qu'un certain nombre de professionnels adhèrent

à la perspective académique de l'évaluation (développement des connaissances) alors que la grande majorité croit que l'évaluation constitue un instrument utile à la prise de décision (Alkin & House, 1992).

Le cadre de réflexion qui guide la réalisation du premier type d'évaluation repose sur l'idée que l'évaluation de programme doit obéir à la même logique et aux mêmes méthodes et techniques que celles associées à la recherche sociale bien qu'elle possède ses propres caractéristiques (Rossi & Wright, 1994). Comme pour la recherche, la composante empirique doit relier les fondements théoriques à la pratique en évaluation de programme. Selon la nature des études empiriques, elles influencent différemment la discipline: (a) Elles peuvent tenter de valider des théories en les appliquant à des cas réels et (b) elles peuvent contribuer au développement (adaptation) du premier niveau de connaissances (noyaux durs et ceintures de défense) grâce au lien intime que l'évaluation entretient avec la pratique.

La réalisation d'une évaluation de ce type nécessite une bonne connaissance et une utilisation juste des théories en évaluation de programme. Ce rapport entre la théorie et la pratique en évaluation de programme a été étudié plus spécifiquement par Christie (2003). Cette étude permet de mieux comprendre comment les théories guident la pratique et comment elles influencent l'évolution de la discipline. Ainsi, les résultats de l'étude montrent qu'une faible proportion des évaluateurs (10%) a l'habitude d'utiliser une théorie particulière dans le cadre de leur pratique (Christie). Bien que certains (36%) connaissent relativement bien une théorie, une grande partie d'entre eux n'emploie pas de modèle théorique dans leur pratique (Christie). De plus, ceux qui affirment utiliser un modèle théorique avouent en modifier certains éléments lorsque vient le moment de l'appliquer. Par conséquent, les pratiques ne correspondent pas directement aux propositions formulées dans les modèles théoriques originaux. Ces résultats appuient les propos tenus par quelques théoriciens (Chelimsky, 1998; Scriven, 1991) qui estiment que la théorie est, pour plusieurs professionnels en évaluation, «un luxe» (Scriven, 1991). En fait, Smith (1993) affirme que souvent les professionnels en évaluation ne se reportent à aucune théorie évaluative pour guider leur travail ou emploient une ou des parties de quelques théories dominantes.

Le second type d'évaluation présente un aspect plus «instrumental» où l'évaluation se distingue tout à fait de la recherche (Borich &

Jemelka, 1982; Guba & Lincoln, 1981; Nadeau, 1988; Worthen & Sanders, 1973). Dans cette optique, l'évaluation répond à des critères additionnels qui la distinguent de la recherche: (a) Sa fonction essentielle est de donner de la rétroaction aux parties intéressées par le programme; (b) son intérêt est de valider et d'expliquer la causalité interne d'un programme; (c) les réalités politiques et administratives du programme qui influencent les décisions quant à la façon de conduire l'évaluation; et (d) ses délais de réalisation, beaucoup plus restreints et rigides que ceux qui marquent les recherches sociales (Rossi & Wright, 1994). Ces différences ont éloigné la pratique de l'évaluation de la recherche et ont influencé les critères qui déterminent sa qualité (Levin-Rozalis, 2000). Selon cette conception, l'évaluation est une source d'informations contribuant principalement à la prise de décision. Ainsi, l'activité évaluative permet de suivre l'évolution d'un programme dans le temps et l'espace, de porter un jugement quant à la poursuite ou non du programme, et de proposer, en cours de réalisation, des modifications lui permettant d'être plus efficace.

Ce travail est utile, mais pour qui? L'évaluation doit-elle répondre aux besoins et attentes des parties prenantes, des gestionnaires, des législateurs, ou des citoyens? Selon l'acteur visé par l'activité, il est à prévoir que la planification, l'exécution, et la publication de l'évaluation seront réalisées différemment. En général, deux types d'évaluation sont reconnus comme utiles à la prise de décision. Le premier type, l'évaluation managériale, aide le gestionnaire et les parties prenantes à prendre de décisions particulières. Le second type, l'évaluation de légitimation, aide les législateurs et les citoyens à mieux comprendre les programmes et à influencer l'intervention publique. Selon Turgeon et Dubois (2000), contrairement à l'évaluation managériale, l'évaluation de légitimation ne limite pas son analyse aux éléments qui composent le processus de gestion de la politique ou du programme. L'évaluation de légitimation s'intéresse à la raison d'être, aux effets, et au rendement des politiques et des programmes. Par ce type d'évaluation, le gouvernement désire juger de la pertinence des programmes, mais aussi mesurer et apprécier leurs effets, leurs répercussions, ainsi que leur rendement (Turgeon & Dubois, 2000).

Au Canada, les évaluations de programme, réalisées autant par les organismes gouvernementaux et les entreprises de consultation que par les chercheurs universitaires, sont majoritairement de nature managériale. Ces études répondent, en grande partie, à des besoins

de gestion et demeurent largement sous le monopole de l'exécutif gouvernemental. Ce type d'évaluation se concentre souvent sur les activités de surveillance ou de suivi d'implantation de programmes qui regroupent un large éventail de données administratives. À la lumière des résultats obtenus par l'évaluation, les administrateurs ajustent leurs activités et leurs façons de faire afin d'améliorer l'efficacité et l'efficience des programmes dont ils ont la responsabilité (Turgeon & Dubois, 2000).

DISCUSSION

L'application de notre modèle d'analyse du développement des disciplines scientifiques conduit à penser que l'évaluation de programme se situe actuellement au troisième stade du développement des disciplines scientifiques qui suppose un changement prochain des fondements dominants de la discipline. Si le modèle d'analyse stipule que la discipline demeure peu de temps au troisième stade de développement en raison de l'instabilité et de la rigueur des débats, l'évaluation semble au contraire avoir atteint un équilibre depuis un long moment et aucun indice ne laisse présumer son retour prochain au deuxième stade de développement.

La structuration des connaissances et la relation entre la théorie et la recherche empirique peuvent expliquer cette situation. Comme pour toute discipline au troisième stade de développement théorique, plus d'un noyau dur cohabite au premier niveau des connaissances en évaluation. Bien que ces deux noyaux durs soient protégés par des ceintures de défense, celles-ci sont beaucoup moins provocantes envers l'autre que ne le suggère la théorie de Kuhn et de Lakatos. Selon notre étude, ces deux milieux sont constitués de théories qui naissent et se développent sans liens évidents avec les autres composantes des ceintures de défense. En principe, l'élaboration des théories devrait être influencé par la présence des théories qui cohabitent dans la même ceinture de défense et par celles qui se développent dans la ceinture de défense adverse.

Plusieurs facteurs peuvent contribuer à la faiblesse des échanges fructueux de points de vue entre les noyaux durs. Premièrement, la nature multidisciplinaire de l'évaluation isolerait les débats théoriques et empiriques dans les disciplines traditionnelles. L'application du modèle d'analyse montre que l'évaluation constitue un système ouvert de connaissances où différents éléments (fondements ontologiques et épistémologiques, concepts théoriques, et données

empiriques) influent sur son développement. Mais l'évaluation de programme gagnerait peut-être à disposer d'une tribune centrale où l'ensemble des académiciens et des professionnels intéressés par l'évaluation pourraient discuter des connaissances relatives à cette discipline. Deuxièmement, il faut constater le faible nombre d'institutions d'enseignement qui offrent des programmes spécialisés en évaluation de programme et qui enseignent l'évaluation de programme comme une discipline académique constituée de fondements théoriques plutôt que comme un simple outil de gestion.

Il se pourrait également que les deux niveaux de connaissances (théorique et pratique) ne soient pas solidement reliés. Il semble d'après certains que, d'une part, les auteurs regroupés autour du noyau dur, au sein de la ceinture de défense, ne développent pas leur argumentation au gré des avancées empiriques et, d'autre part, les praticiens n'utilisent pas les théories situées dans les ceintures de défense. Plusieurs auteurs, dont Scriven (1998) et Shadish (1998), ont soulevé ce problème et l'étude de Christie (2003) tendrait à le confirmer. Les évaluations seraient souvent réalisées sans que des liens soient clairement établis avec les théories de l'évaluation et seraient surtout influencées par les expériences et les idées des praticiens quant à ce que doit être l'évaluation. Différents facteurs expliqueraient cette situation. Le caractère intrinsèquement utilitaire accordé à l'évaluation de programme en serait un premier. Pour plusieurs, la discipline est pragmatique et doit répondre aux besoins du client plutôt que de chercher à valider les connaissances théoriques de la discipline (King, 2003). Un deuxième facteur serait lié aux demandes trop pressantes des gouvernements à intégrer la fonction évaluative dans la gestion courante des programmes où les contextes politiques et les délais de réalisation deviennent très contraignants (King, 2003). Troisièmement, le manque d'appui financier pour la recherche empirique en évaluation de programme conduirait trop souvent les praticiens à se concentrer à la réalisation d'évaluations commanditées, de type managérial. Quatrièmement, il est évoqué que le nombre élevé «d'experts accidentels» qui exercent la fonction principale d'évaluateur de programme alors qu'ils sont diplômés d'une discipline connexe à l'évaluation (Christie) serait un facteur non négligeable. Souvent, ces personnes ne possèderaient que peu ou pas de notions relatives à la discipline évaluative (Nadeau, 1988).

Théoriquement, ces deux niveaux de connaissances devraient être en interaction continue permettant ainsi un développement conjoint des deux niveaux de connaissance. La faiblesse des liens unis-

sant les ceintures de défense (théories) et la pratique risque fort d'influencer de façon notable l'évolution de la discipline. Ainsi, en l'absence de lien entre les deux niveaux de connaissances, la validation de théories n'est guère possible, ce qui oblige l'évaluation à demeurer éternellement au troisième stade de développement de la discipline. De plus, sans lien avec la pratique, la théorie risque de ne pas perdurer, car elle demeure un ensemble de concepts sans fondements et sans intérêts pour la discipline. Finalement, sans ce lien entre les deux niveaux de connaissances, l'évaluation risque d'être plus sensible aux exigences (attentes) politiques et administratives qu'au développement à long terme de la discipline. Cet état de fait peut conduire l'évaluation à une critique plus virulente et à de nombreux désaccords quant à sa réalisation et à sa contribution à la prise de décision (Scriven, 1996, 1998). Pour Lakatos, l'absence de relation entre la théorie et la recherche empirique doit être jugé sévèrement: ceci conduit la discipline vers une dégénérescence de ses fondements théoriques.

Les conclusions de cette analyse de la discipline reposent sur les éléments du modèle inspiré des idées de Kuhn et Lakatos. Ces idées sur l'évolution des paradigmes et des programmes de recherche sont nées à une époque où les grands débats scientifiques ont porté sur le choix des modèles de représentation de la réalité, surtout en sciences physiques. Le fait que ces analyses des disciplines scientifiques n'aient pas eues, à l'époque, à se confronter à la remise en question radicale de la science par la pensée constructiviste peut limiter dangereusement notre propre analyse de l'évaluation. Aussi faut-il prendre soin en dernière instance d'examiner l'évaluation avec un peu de recul, en considérant sommairement les liens que cette discipline entretient avec le domaine plus large des sciences sociales.

L'évaluation de programme puise généreusement, on l'a vu, dans l'ensemble des concepts et des méthodologies des sciences sociales. La discipline de l'évaluation ne peut pas ne pas ressentir profondément les tourments vécus dans tout ce vaste domaine scientifique. Or les sciences sociales vivent actuellement les mêmes divisions observées en évaluation entre post-positivistes et constructivistes. L'évaluation ne serait-elle pas également à la remorque d'un plus vaste mouvement de remise en question épistémologique, qui va bien au-delà de la théorie au sens étroit du terme?

Car, indéniablement, l'usage du mot théorie par les plus grands penseurs en évaluation déborde de beaucoup le sens habituellement prêté

à ce terme qui essentiellement se reporte à la notion de modèle de représentation de la réalité. La «théorie» de l'évaluation, au contraire, engloberait, d'après la majorité des auteurs en évaluation, l'ensemble de ses fondements, de ses concepts, de ses approches à la connaissance, et, enfin, de ses modèles de représentation de la réalité (Alkin, 2003). Les révolutions scientifiques, habituellement de courte durée, renvoient aux divisions sur les théories au sens étroit du terme et non au sens large comme le veut plutôt la tradition en évaluation.

N'est-ce pas la clé qui permet d'expliquer cette révolution scientifique qui n'en finit pas? L'ensemble des sciences sociales vit une division profonde dans la réflexion et dans la pratique relativement à l'approche de la connaissance. De toute évidence, c'est beaucoup plus profond, beaucoup plus radical comme opposition qu'un débat entre deux modèles de représentation de la réalité qui partagent, dans les visions de Kuhn et Lakatos, la même conception de la science. La réconciliation est prévue à terme dans leurs visions parce que les scientifiques sont sensés partager le même objectif, celui de bâtir un modèle de représentation de la réalité. C'est une vision ancrée profondément dans le positivisme logique qui ne prévoyait sans doute pas, à l'époque, le conflit profond qui sévit actuellement entre le post-positivisme et le constructivisme.

Que partagent réellement aujourd'hui les post-positivistes et les constructivistes en sciences sociales comme en évaluation? Ce partage suffit-il à les forcer, à terme, à la réconciliation? La révolution scientifique qui ne doit s'échelonner que sur une certaine période, selon Kuhn, lorsqu'elle met aux prises des scientifiques qui partagent la même conception de la science, peut-elle perdurer si ceux qui participent à l'évolution de la discipline de l'évaluation s'opposent radicalement sur la façon même de faire de la science? L'évaluation de programme n'est-elle pas à la remorque d'un débat profond sur la conception de la science qui l'entraîne dans une révolution qui n'en finit pas? Le schisme épistémologique qui sévit en sciences sociales ne se reflète-t-il pas totalement en évaluation et ne faut-il pas plutôt envisager une longue cohabitation de pratiques scientifiques et évaluatives profondément désarticulées?

L'évolution de l'évaluation de programme, dans un tel contexte, se fera peut-être à l'intérieur de plus vastes mouvements de la pensée scientifique, et la tolérance autant que le pragmatisme faciliteront peut-être ce long passage obligé. Faudra-t-il pour autant négliger la construction de noyaux durs et de ceintures de défense cohérents et

empiriquement puissants? Ou si, au contraire, cette pratique se devra d'être renforcée pour mieux résister au relativisme qui ne manquera pas de s'installer au sein des penseurs et praticiens?

En conclusion, il nous semble que si cette dynamique de révolution scientifique qui perdure est bien réelle et que l'évaluation traverse véritablement cette révolution en synchronie avec l'ensemble des sciences sociales, force alors est de constater que l'évaluation n'est ni en avance, ni en retard quant à son évolution par rapport aux autres sciences sociales.

NOTES

1. La première version de cet article a reçu le premier prix/mention honorable dans la Division des étudiants diplômés dans le cadre du concours de dissertation étudiante de la Société canadienne d'évaluation.
2. Toutefois, les auteurs précisent que certains éléments peuvent être partagés par plusieurs personnes, voire par les membres d'une société entière. De plus, l'évaluateur, les intervenants, et l'objet évalué sont en interaction continue et ils s'influencent mutuellement (Greene, 2000; Guba & Lincoln, 1989; Patton, 1997a, 1997b).

RÉFÉRENCES

- Alkin, M. (2003). Evaluation theory and practice: Insights and new directions. Dans C. Christie (Éd.), *The practice-theory relationship in evaluation* (New Directions for Program Evaluation, N° 97, pp. 81–90). San Francisco: Jossey-Bass.
- Alkin, M., & House, E. (1992). Evaluation of programs. Dans M. Alkin (Éd.), *Encyclopedia of education* (pp. 463–467). Toronto: American Educational Research Association, Maxwell Macmillan Canada.
- Bernstein, I., & Freeman, H. (1975). *Academic and entrepreneurial research*. New York: Russell Sage.
- Bickman, L. (1990). Using program theory to describe and measure program quality. Dans L. Bickman (Éd.), *Advances in program theory* (New Directions for Program Evaluation). San Francisco: Jossey-Bass.

- Bickman, L. (1994). An optimistic view of evaluation. *Evaluation Practice*, 15(3), 255–259.
- Borich, G., & Jemelka, R. (1982). *Programs and systems: An evaluation perspective*. New York: Academic Press.
- Campbell, D. (1975). Degrees of freedom and the case study. *Comparative Political Studies*, 8, 178–193.
- Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- Chalmers, A. (1987). *Qu'est-ce que la science? Popper, Kuhn, Lakatos, Feyerabend*. Paris: La découverte.
- Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*, 19(3), 35–55.
- Chen, H. (1990). *Theory driven evaluations*. Newbury Park, CA: Sage.
- Chen, H., & Rossi, P. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces*, 59, 106–122.
- Chen, H., & Rossi, P. (1989). Issues in the theory-driven perspective. *Evaluation and Program Planning*, 12(4), 299–306.
- Christie, C. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. Dans C. Christie (Éd.), *The practice-theory relationship in evaluation* (New Directions for Program Evaluation, N° 97, pp. 7–36). San Francisco: Jossey-Bass.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cronbach, L. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672–683.
- Cronbach, L. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L., & Snow, R. (1969). *Individual differences in learning ability as a function of instructional variables*. Stanford, CA: Stanford University Press.

- Eisner, E. (1975). *The perceptive eye: Toward a reformation of educational evaluation*. Invited address, Division B: Curriculum and Objectives, American Educational Research Association, Washington, DC.
- Eisner, E. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation. Dans G. Madaus, M. Scriven, & D. Stufflebeam (Éds.), *Evaluation models*. Boston: Kluwer-Nijhoff.
- Flinders, D., & Eisner, E. (2000). Educational criticism as a form of qualitative inquiry. Dans D. Stufflebeam et al. (Éds.), *Evaluation models*. Boston: Kluwer.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldines.
- Glass, G. (1975). A paradox about excellence of schools and the people in them. *Educational Researcher*, 4, 9–13.
- Greene, J. (2000). Understanding social programs through evaluation. Dans N. Denzin & Y. Lincoln (Éds.), *Handbook of qualitative research* (2e éd.). Thousand Oaks, CA: Sage.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation*. San Francisco: Jossey-Bass.
- Guba, E., & Lincoln, Y. (1982). Epistemological and methodological bases of naturalistic inquiry. *Education Communication and Technology Journal*, 30(4), 233–252.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*, Newbury Park, CA: Sage.
- House, E., & Howe, K. (1998). *Deliberative democratic evaluation in practice*. Boulder, CO: University of Colorado.
- Kee, J. (1995). Benefit-cost analysis in program evaluation. Dans J. Wholey et al. (Éds.), *Handbook of practical program evaluation* (pp. 456–488). San Francisco: Jossey-Bass.
- Kidder, L., & Fine, M. (1987). *Qualitative and quantitative methods: When stories converge. Multiple methods in program evaluation* (New Directions for Program Evaluation). San Francisco: Jossey-Bass.

- King, J. (2003). The challenge of studying evaluation theory. Dans C. Christie (Éd.), *The practice-theory relationship in evaluation* (New Directions for Program Evaluation, N° 97, pp. 57–68). San Francisco: Jossey-Bass.
- Kuhn, T. (1983). *La structuration des révolutions scientifiques* (L. Meyer. Trad.). [*The structure of scientific revolutions*] (Originel publié 1962). Paris: Flammarion.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. Dans I. Lakatos & A. Musgrave (Éds.), *Criticism and the growth of knowledge*. Proceedings of the International Colloquium in the Philosophy of Science, London, 1965 (vol. 4). Cambridge: Cambridge University Press.
- Levin, H. (1983). *Cost-effectiveness: A primer* (New Perspectives in Evaluation, 4). Newbury Park, CA: Sage.
- Levin-Rozalis, M. (2000). Abduction: A logical criterion for programme and project evaluation. *Evaluation*, 6(4), 415–432.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lindquist, E. (1953). *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin.
- Madaus, G., Scriven, M., & Stufflebeam, D. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff.
- Mills, C.W. (1967). *L'imagination sociologique*. Paris: Maspero.
- Miron, G. (1998). Choosing the right research methods: Qualitative? Quantitative? Or both? Dans L. Buchert (Éd.), *Education reform in the south in the 1990s*. Paris: Unesco.
- Nadeau, M-A. (1988). *L'évaluation de programme: Théorie et pratique* (2e éd.). Sainte-Foy, QC: Presses de l'Université Laval.
- Patton, M. (1990). *Qualitative evaluation and research methods* (2e éd.). Newbury Park, CA: Sage.
- Patton, M. (1997a). Toward distinguishing empowerment evaluation and placing it in a large context. *Evaluation Practice*, 18(2), 147–163.

- Patton, M. (1997b). *Utilization-focused evaluation: The new century text* (3e éd.). Thousand Oaks, CA: Sage.
- Patton, M. (2000). Utilization-focused evaluation. Dans D. Stufflebeam et al. (Éds.), *Evaluation models*. Boston: Kluwer.
- Platt, J. (1992). Case study in American methodological thought. *Current Sociology*, 40(1), 17–48.
- Popper, K.R. (1991). *La connaissance objective*. Paris: Aubier.
- Rogers, P. (2000). Program theory: Not whether programs work but how they work. Dans G. Stufflebeam et al. (Éds.), *Evaluation models*. Boston: Kluwer.
- Rossi, P., & Freeman, H. (1985). *Evaluation: A systematic approach* (3e éd.). Thousand Oaks, CA: Sage.
- Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation: A systematic approach* (6e éd.). Thousand Oaks, CA: Sage.
- Rossi, P., & Wright, S. (1994). L'évaluation: Appréciation de la théorie, de la pratique et des aspects politiques. *Revue Politiques et Management Public*, 12(4), 93–135.
- Schatzman, L., & Strauss, A. (1973). *Field research*. Englewood Cliffs, NJ: Prentice-Hall.
- Scriven, M. (1967). The methodology of evaluation. Dans R. Stake (Éd.), *Curriculum evaluation (AERA Monograph Series on Curriculum Evaluation, 1)*. Chicago: Rand McNally.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20(1), 147–166.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17(2), 151–161.

- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19(1), 57–70.
- Shadish, W. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1–19.
- Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation: Theories of practice*. London: Sage.
- Shadish, W., & Epstein, R. (1987). Patterns of program evaluation among members of the Evaluation Research Society and Evaluation Network. *Evaluation Review*, 11, 555–590.
- Smith, L., & Pohland, P. (1974). Educational technology and the rural highlands. Dans L. Smith (Éd.), *Four examples: Economic, anthropological, narrative, and portrayal* (AERA Monograph on Curriculum Evaluation). Chicago: Rand McNally.
- Smith, N. (1993). Improving evaluation theory through the empirical studies of evaluation practice. *Evaluation Practice*, 14(3), 237–242.
- Stake, R. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523–540.
- Stake, R. (1975). *Program evaluation: Particular responsive evaluation*. Western Michigan University Evaluation Center, Paper no. 5.
- Stake, R., Easley, J., & Anastasiou, K. (1978). *Case studies in science education*. Washington, DC: National Science Foundation, Directorate for Science Education, Office of Program Integration.
- Stufflebeam, D. (1966). A depth study of the evaluation requirement. *Theory Into Practice*, 5, 121–134.
- Stufflebeam, D. (1967). The use of and abuse of evaluation in Title III. *Theory Into Practice*, 6, 126–133.
- Stufflebeam, D. (2001). *Evaluation models* (New Directions for Program Evaluation, N° 89). San Francisco: Jossey-Bass.
- Suchman, E. (1967). *Evaluative research*. New York: Russell Sage Foundation.

- Tsang, M. (1997). Cost analysis for improved educational policy making and evaluation. *Educational Evaluation and Policy Analysis*, 19(4), 318–324.
- Turgeon, J., & Dubois, N. (2000). L'évaluation des politiques et des programmes au niveau du gouvernement fédéral canadien. Dans Conseil national de l'évaluation (Éd.), *L'évaluation au service de l'avenir: Rapport annuel 1999* (pp. 183–204). Paris: La documentation française.
- Tyler, R. (1932). *Service studies in higher education*. Columbus, OH: Bureau of Educational Research, Ohio State University.
- Vallance, E. (1973). *Aesthetic criticism and curriculum description*. Ph.D. dissertation, Stanford University.
- Weiss, C. (1972). *Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Weiss, C. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis*, 3, 531–546.
- Weiss, C. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives, for children and families. Dans A. Connell et al. (Éds.), *New approaches to evaluating community initiatives*. New York: Aspen Institute.
- Weiss, C. (1997). *Theory-based evaluation: Past, present, and future*. *American Evaluation Association: Progress and future directions in evaluation* (New Directions for Program Evaluation, N° 76, pp. 41–55). San Francisco: Jossey-Bass.
- Weiss, C., & Buccuvalas, M. (1980). Truth tests and utility tests: Decision-makers' frames of reference for social science research. *American Sociological Review*, 45, 302–313.
- Wholey, J. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute.
- Worthen, B., & Sanders, J. (1973). *Educational evaluation: Theory and practice*. Worthington, OH: Jones.



Nathalie Dubois est étudiante au doctorat en analyse et management des politiques publiques à l'École nationale d'administration publique.

Richard Marceau est professeur titulaire à l'École nationale d'administration publique.

