

## JUGEMENT CRÉDIBLE EN ÉVALUATION DE PROGRAMME : DÉFINITION ET CONDITIONS REQUISES

Marthe Hurteau  
Université du Québec à Montréal  
Montréal, Québec

Pierre Valois  
Université du Luxembourg  
Walferdange, Luxembourg

Amélie Bossiroy  
Université du Québec à Montréal  
Montréal, Québec

**Résumé :** L'évaluation de programme consiste à établir la qualité d'un programme afin de générer un jugement qui doit être valide et crédible aux yeux des détenteurs d'enjeux. La présente recherche vise à mieux comprendre le terme « jugement crédible » et à cerner les conditions requises pour le générer. Des entrevues semi-dirigées ont été réalisées auprès de six évaluateurs chevronnés et clients. Les résultats mettent en évidence la présence des conditions afin d'assurer un jugement crédible : une méthodologie rigoureuse, une information de qualité, la représentation de différents points de vue, et une cohérence dans le développement de l'argumentation. La contribution de cette étude réside dans le fait qu'elle valide la réflexion de certains théoriciens, tout en soulignant les embûches qui se présentent au cours du passage entre la théorie et l'exercice professionnel.

**Abstract:** Program evaluation aims at establishing the quality of a program in order to generate a judgment that is valid and credible for the stakeholders. The present research aimed to gain a better understanding of the term “credible judgment” and to identify the required conditions to generate it. Semi-structured interviews were conducted with six experienced evaluators and clients. The results highlight the required conditions to ensure credible judgment: rigorous methodology, quality information, representation of various viewpoints, and coherence in developing argumentation. The contribution of this study consists in validating the

---

Correspondance à l'auteure : Marthe Hurteau, Département d'éducation et pédagogie, Faculté des sciences de l'éducation, Université du Québec à Montréal; <hurteau.marthe@uqam.ca>

thinking of particular theorists, while pointing out the pitfalls that arise in the transition from theory to professional practice.

■ L'évaluation de programme se définit comme une cueillette systématique d'informations sur les différentes composantes d'un programme ainsi que leurs interactions, et ce, dans le but d'en faire une description détaillée, d'en comprendre le fonctionnement et, ultimement, de porter un jugement (Chen, 2005; Fitzpatrick, Sanders, & Worthen, 2011; Guba & Lincoln, 1981; Mathison, 2005; Nevo, 2006; Patton, 1997; Scriven, 1991). Les recommandations qui en découlent devraient être en mesure de soutenir les éventuelles prises de décision concernant les orientations futures du programme (Fournier & Smith, 1993; House, 1980; Mathison; Weiss, 1998). Cette démarche se base sur la prémisse que le jugement doit être à la fois valide sur le plan scientifique et « acceptable » (*believable*) aux yeux des détenteurs d'enjeux (*stakeholders*) (Schwandt, 2009). Greene (2006) identifie quatre types de détenteurs d'enjeux, soit : les administrateurs d'un programme, ses gestionnaires, ses clients, ainsi que ceux qui sont indirectement affectés par le programme. Scriven (2007) complète la liste avec les regroupements politiques.

Depuis ses débuts, l'évaluation de programme a été rapidement confrontée à de sérieuses remises en question quant à sa capacité de générer un jugement valide. Dès 1972, dans son article intitulé *The Failure of Educational Evaluation*, Guba émet une hypothèse d'explication à l'effet que les évaluateurs éprouvent des difficultés à sélectionner les critères requis pour générer un jugement fondé et à les documenter. Dans son article, Scriven (1995) appuie les conclusions de son collègue. Fournier et Smith (1993), quant à eux, soulignent que les évaluateurs accordent trop souvent d'importance à justifier les conclusions, sans préalablement s'assurer qu'elles sont crédibles aux yeux des détenteurs d'enjeux. Les résultats de la méta-analyse réalisée par le Secrétariat du Conseil du Trésor du Canada (2004) confirment ces observations. L'étude porte sur l'analyse de 115 rapports d'évaluation et même si Daigneault (2010) identifie plusieurs lacunes importantes qui entachent la crédibilité des conclusions, l'auteur considère qu'elle demeure pertinente et intéressante. Cela est d'autant plus vrai que ce genre d'étude est peu courant. Ainsi, il en ressort que seulement 21 % des rapports génèrent des conclusions fondées sur des critères explicites. De plus, si l'on considère que, d'une part, 34 % des rapports n'ont pu être analysés parce qu'ils ne fournissaient pas l'information requise et que cela constitue un problème en soi, et que, d'autre part, les résultats ont été générés à partir d'un échantillon de rapports biaisé,

Daigneault est justifié de considérer la teneur des résultats comme étant plus optimiste qu'autre chose.

Au cours des 30 dernières années, des théoriciens ont développé plusieurs modèles afin de baliser la démarche évaluative et favoriser la production d'un jugement valide. Cependant, diverses études, pour ne mentionner que celles de Stake et al. (1997), de Christie (2003), de Arens (2006), ainsi que de Hurteau, Houle, et Mongiat (2009), concluent que ces modèles s'avèrent peu utiles, possiblement parce qu'ils ne traduisent pas les façons de faire observées chez les praticiens. Christie conclut d'ailleurs son étude par la question suivante : « if formal theory does not guide their work, then what does? » (p. 92). Les récents travaux de Donaldson, Christie, et Mark (2009) portent sur les modalités requises pour générer des données probantes crédibles. Cependant, Schwandt (2009) note que « developing credible evidence in evaluation is not sufficient for establishing the credibility of an evaluation » (p. 209).

Mais quelles sont ces conditions requises? La présente démarche vise à cerner le concept de jugement crédible dans le cadre de l'évaluation de programme ainsi que les conditions requises pour le générer. Dans un premier temps, nous définirons le concept général de jugement dans le cadre de l'évaluation de programme, de même que celui de jugement crédible. Par la suite, nous présenterons la méthodologie retenue aux fins de la présente étude, les résultats générés, ainsi que les conclusions qui en découlent.

## LE CADRE CONCEPTUEL

### Le positionnement du jugement en évaluation de programme

À son origine, l'évaluation était identifiée comme l'activité permettant d'établir la valeur d'un programme (Scriven, 1967). Cet auteur en vient à proposer le *processus logique sous-jacent à l'évaluation* (notre traduction du terme « *the logic of evaluation* ») (Scriven, 1980), qui requiert quatre opérations, soit : (a) l'établissement des critères, (b) la construction des seuils de performance ou des standards, (c) la mesure de la performance et sa comparaison aux seuils, et (d) la synthèse et l'intégration des données afin de générer un jugement. Par la suite, des chercheurs s'en sont inspirés en partie ou en totalité pour développer d'autres modèles (e.g., Arens, 2006; Fish & Coles, 1998; Fournier, 1995; Hurteau, Lachapelle, & Houle, 2006; Kundin, 2010; Stake, 2004). Il est à noter que ces modèles ont été élaborés à partir de connaissances théoriques, c'est-à-dire de la synthèse de contributions antérieures.

Avec le temps, les visées de la démarche évaluative ont élargi leur spectre. Ainsi, Talmage (1982) envisage l'évaluation comme un soutien à la prise de décision pour ceux qui élaborent des politiques. Preskill et Torres (1998) et Mark, Henry, et Julne (2000) prônent une évaluation à tendance formative afin d'améliorer les programmes. Dans la même veine, Patton (1994, 2005) propose *l'évaluation développementale* qui offre une rétroaction continue afin de permettre aux organisations d'apporter les ajustements requis. Preskill et Torres (2000) ainsi que King (2002) appuient cette perspective. Weiss (1998) et Henry (2000), quant à eux, considèrent que l'évaluation devrait réduire les problèmes sociaux et satisfaire certains besoins humains. Finalement, House et Howe (1999) suggèrent qu'elle donne une voix aux plus démunis afin d'assurer une justice sociale et une équité. Cette perspective est reprise par Greene (2006) avec le concept *d'évaluation démocratique*. Ces différents points de vue complémentaires permettent de mieux répondre aux aspirations des différentes personnes concernées. La typologie d'Alkin et Christie (2004) propose trois critères pour regrouper 28 modèles d'évaluation, soit le *jugement*, la *méthodologie* et l'*utilisation*. Plus récemment, Fitzpatrick, Sanders, et Worthen (2011) proposent une typologie comprenant quatre catégories pour classer les modèles, soit : le *jugement éclairant la qualité d'un programme ou d'un produit*, l'*amélioration du programme*, les *décisions devant être envisagées*, et la *participation des détenteurs d'enjeux*.

Il ressort de ce bref tour d'horizon que même si la production d'un jugement valide n'a pas constitué l'unique visée de l'évaluation de programme, il n'en demeure pas moins qu'elle a constitué une option valable à travers le temps.

### Le jugement crédible en évaluation de programme

Dès 1980, House avance que, pour être crédible, le jugement doit rencontrer certaines conditions, à savoir : (a) faire preuve d'une *cohérence interne*, c'est-à-dire que la relation entre les visées de la démarche, les données recueillies, et le jugement émis doit ressortir de façon évidente; (b) faire preuve d'esthétisme (façon de le présenter); (c) mobiliser l'ensemble des personnes concernées; et (d) être authentique dans le sens qu'il constitue le reflet de la réalité. Lincoln et Guba (1985) reprennent cette dernière condition qu'ils associent à un de leurs quatre critères pour assurer la rigueur de la démarche évaluative, soit la *crédibilité*. Le jugement est ainsi envisagé comme le fruit d'une négociation entre l'évaluateur et les différents détenteurs d'enjeux dans le but de dégager une vision

intégrée du programme dont la facture se rapproche le plus possible de la réalité. Chen (2004), quant à lui, associe la crédibilité à la combinaison entre la validité scientifique et une reconnaissance de la part des détenteurs d'enjeux (*stakeholders*) qui s'attendent à ce que la démarche soit juste, pertinente, et utile. Greene (2006) complète cette idée en soulignant que cette reconnaissance doit *se mériter*. Pour y parvenir, l'évaluateur doit valoriser une démarche inclusive, un dialogue ininterrompu, ainsi qu'une interaction dynamique avec les détenteurs d'enjeux.

En résumé, pour être crédible, le jugement devra non seulement revêtir une validité scientifique, mais il doit aussi être crédible (believable) aux yeux des détenteurs d'enjeux (stakeholders). En plus d'épouser le point de vue de Schwandt (2009), ce positionnement permet ainsi aux détenteurs d'enjeux de : « construire, individuellement ou collectivement, un jugement susceptible de se traduire en actions » (Champagne, Contandriopoulos, Brousselle, Hartz, & Denis, 2009, p. 38). Finalement, il faut retenir que ces différents points de vue émergent d'une *connaissance théorique* et non *pratique* (Donaldson & Lipsey, 2008; Schwandt, 2007).

## OBJECTIF DE LA RECHERCHE

La présente étude vise à explorer et analyser les modalités pour produire un jugement crédible dans le cadre de l'évaluation de programme en adoptant une perspective axée sur la « connaissance pratique » afin de rendre explicite les éléments de réponse à notre questionnement (Schön, 1983), à savoir, qu'est-ce qu'un jugement crédible et quelles conditions requiert-il? Il ne s'agit donc pas de vérifier des théories existantes comme de tenter de jeter un regard neuf sur des phénomènes déjà étudiés.

## MÉTHODOLOGIE

Les visées de la présente recherche s'inscrivent dans l'optique de la *Grounded Theory* (Guillemette, 2006) qui se caractérise par la suspension temporaire et volontaire du recours à des cadres théoriques existants au profit d'une ouverture à ce qui émerge des données de terrain, par sa façon particulière de préciser l'objet de recherche, par une interaction circulaire entre la collecte et l'analyse des données, et par les procédures d'analyse des données favorisant une ouverture à l'émergence.

## Échantillonnage

Un échantillon intentionnel, à variation maximale, a été retenu aux fins de la présente recherche (Given, 2008). Six sujets ont été choisis en fonction de leur capacité à nous aider à mieux comprendre le phénomène (jugement crédible) et non pas seulement à le décrire. Il y a lieu de croire que différentes perspectives sont représentées puisque trois d'entre eux sont considérés comme des clients de l'évaluation, soit les personnes à qui les rapports d'évaluation sont destinés, alors que les trois autres sont des évaluateurs chevronnés, soit les personnes qui réalisent l'évaluation et produisent des rapports. Il n'existe cependant pas d'appariement direct entre les deux groupes. De plus, trois répondants provenaient d'organismes gouvernementaux fédéraux (mandats de grande envergure) alors que les trois autres provenaient d'organismes communautaires (mandats de nature plus locale). Tous occupaient des fonctions de cadre intermédiaire ou supérieur et ils étaient détenteurs d'une maîtrise. Pour des raisons de confidentialité, il n'est pas possible de développer davantage sur la nature des organisations concernées ainsi que les postes que ces personnes occupent.

## Méthodes

Afin d'explorer la thématique, nous avons choisi de réaliser des entretiens semi-structurés (60 à 90 minutes) en s'appuyant sur un questionnaire. Tout comme Tourmen (2009), nous avons mis l'accent sur les façons de faire des répondants afin de faire ressortir les connaissances pratiques non théoriques. Cependant, alors que cette auteure s'est appuyée sur la planification d'une évaluation, nous avons plutôt invité les répondants à se référer à des rapports d'évaluation qu'ils avaient rédigés ou étudiés, selon le cas. Les questions abordaient le contenu d'un rapport (présence et pondération des éléments) ainsi que le contenant (traitement et présentation des éléments). Certaines des questions qui leur ont été posées seront présentées plus loin dans le texte. Leur teneur et leur ordre de présentation pouvaient différer sensiblement d'un répondant à une autre afin d'optimiser une conceptualisation et une théorisation émergente (Charmaz, 2002). Finalement, l'analyse progressive des résultats a été prise en considération dans les entretiens subséquents afin d'enrichir la réflexion, tout en préservant la perspective d'ouverture à l'émergence (Charmaz, 2005).

## Déroulement

Les participants recevaient à l'avance le formulaire de consentement ainsi que le questionnaire afin qu'ils puissent se préparer à l'entrevue.

Ces documents les informaient sur les objectifs de la recherche, notre souhait d'enregistrer l'entrevue, et nos devoirs et obligations à leur égard en tant que chercheurs. Compte tenu de son expérience en la matière, les entrevues ont été réalisées par la chercheure principale. Cela était particulièrement important dans le contexte actuel puisqu'il s'agissait d'accompagner les répondants dans leur réflexion plutôt que de leur demander de répondre systématiquement à une série de questions. Tous ont consenti à ce que les entretiens soient enregistrés et textuellement retranscrits.

### Modalités d'analyse des données

Toujours dans le but de s'inscrire dans l'optique de la Grounded Theory, l'analyse des données a été réalisée dans une perspective essentiellement inductive afin de favoriser l'émergence des connaissances et des compréhensions à partir du discours des répondants. L'encodage des données et leur analyse ont été effectués à l'aide du logiciel QDA Miner. Ces opérations ont été effectuées séparément par la chercheure et son assistante. Elles ont suivi les étapes proposées par Blais et Martineau (2006), soit la préparation des données brutes, une lecture attentive et approfondie, l'identification et la description des premières catégories, et la révision et le raffinement des catégories. C'est ainsi que le choix final des catégories émane d'un processus de triangulation des points de vue.

## RÉSULTATS

La présente section traite de la définition d'un jugement crédible et des conditions requises pour y parvenir. D'entrée de jeu, soulignons qu'il se dégage un consensus quant aux points de vue émis par les répondants, et ce, quel que soit leur groupe d'appartenance.

### Qu'est-ce qu'un jugement professionnel crédible?

D'emblée, les répondants se sentent à l'aise avec l'idée que le jugement constitue le produit d'une évaluation. Pour être *crédible*, le jugement devra cependant être « argumenté, utile, et explicite ». Par *argumenté*, les répondants réfèrent au fait qu'il doit s'appuyer sur de l'information de qualité (contenu) présentée adéquatement (contenant). Par *utile*, ils réfèrent au fait que les jugements devraient permettre d'envisager des suites. Ils précisent que l'évaluateur devrait prendre en considération les éventuelles répercussions (tant positives que négatives) et faire

preuve de pédagogie en présentant les résultats de façon constructive. Finalement, par *explicite*, ils réfèrent à un jugement qui est clair et catégorique, dans lequel ils ne retrouvent aucune ambiguïté. Ainsi, l'évaluateur ne devrait laisser aucune question en suspens quant à la teneur du jugement.

### Les conditions requises

Tel qu'indiqué dans le tableau 1, la majorité des répondants rapportent que le jugement est *crédible* s'il (a) s'appuie sur une méthodologie rigoureuse, (b) repose sur une information de qualité, (c) prend en considération les différents points de vue des personnes concernées, et (d) développe une argumentation cohérente. Chacune de ces conditions est développée dans les paragraphes qui suivent.

**Tableau 1**  
**Opinions des répondants au regard des conditions requises pour générer un jugement crédible**

<i>Conditions requises</i>	<i>Proportions</i>
Méthodologie rigoureuse	6/6
Qualité de l'information générée	6/6
Représentation des différents points de vue	6/6
Cohérence du développement	5/6

#### Condition 1 : s'appuyer sur une méthodologie rigoureuse

Pour les répondants, la méthodologie doit être inattaquable. À la question « Si vous n'aviez que quelques minutes pour prendre connaissance d'un rapport d'évaluation, que regarderiez-vous? », tous mentionnent la méthodologie, sans hésitation. Ils font référence à la sélection des répondants, aux instruments de mesure, et aux modalités de collecte et d'analyse des données. La présentation des limites de l'évaluation est associée à cette considération et il est évident que les répondants ne la souhaitent pas trop développée. De plus, un répondant précise qu'il apprécie aussi retrouver les critères et les seuils ayant permis d'émettre le jugement. En clair, en s'inscrivant dans la tradition de la recherche scientifique, la méthodologie assure une légitimité à la démarche d'évaluation et lui évite de prêter le flanc à la critique. Aux questions : « Si vous aviez à présenter un rapport auprès d'une



instance, quels éléments devraient être présents afin que vous vous sentiez tout à fait à l'aise parce que vous sentez le jugement fondé? » et « Au contraire, quels éléments (présents ou absents) vous empêcheraient de vous sentir en confiance ou d'être à l'aise? », les répondants rapportent encore la méthodologie et justifient leur choix dans le fait que c'est la première cible qui prête le flanc aux attaques et critiques des clients lorsqu'ils sont insatisfaits des résultats générés par l'évaluation. Cependant, tous reconnaissent que si la méthodologie est essentielle, elle n'en demeure pas moins insuffisante, seule, pour assurer la crédibilité du jugement.

### Condition 2 : reposer sur une information de qualité

Pour les répondants, un jugement crédible devrait reposer sur deux types d'information. Le premier type regroupe les informations « prévues », soit l'ensemble des mesures et des appréciations qui ont été prévues afin de documenter la question d'évaluation. Le second type d'information émerge en cours de route. Il relève de l'intuition de l'évaluateur ainsi que de sa connaissance raffinée du contexte local. Pour les répondants, « c'est comme si l'évaluateur possédait de longues antennes très aiguës et très sensibles qui lui permettent d'aller chercher ces nouveaux éléments ». Ils considèrent cette information essentielle parce qu'elle les rassure quant au fait que l'évaluateur a vraiment cerné les enjeux importants et qu'il a ainsi accompli son mandat. Le témoignage qui suit, tiré de Bossiroy (2009), résume ce point de vue. Je pense à certaines évaluations qui étaient merveilleuses sur le plan méthodologique... tout était parfait! Cependant, il n'y avait rien de nouveau qui se dégageait du rapport et je n'étais pas persuadée que les évaluateurs avaient vraiment saisi le contexte du programme ou qu'ils avaient compris son environnement... je n'avais pas l'impression qu'ils avaient réellement saisi l'ensemble des enjeux. (p. 59, notre traduction)

La proportion des deux types d'information varie selon les caractéristiques du programme (taille et nature), la nature du mandat d'évaluation (e.g., reddition de comptes par rapport à l'évaluation effectuée dans une perspective de bonification, d'ajustement du programme), ainsi que les caractéristiques de l'évaluateur (e.g., expérience, compétence, et personnalité). Finalement, s'ils peuvent toujours « vivre avec le fait » qu'une évaluation ne génère pas des données empiriques, il leur apparaît impensable que les informations prévues soient absentes. En effet, elles sont requises, indispensables, puisque liées à la question d'évaluation et qu'elles constituent par le fait même le cœur de l'évaluation.

Condition 3 : prendre en considération les différents points de vue

Pour les répondants, il est indéniable que le point de vue de l'ensemble des détenteurs d'enjeux doit être pris en considération de façon à dresser un portrait fidèle de la réalité. Ces derniers sont considérés comme des sources d'information complémentaires. Les répondants avouent parfois éprouver une certaine crainte au moment du dépôt du rapport. En effet, ils ont peur que sa teneur suscite un boycottage ou un rejet de la part de certains détenteurs d'enjeux qui auront l'impression que les différents points de vue ne sont pas « bien représentés ». Tout en comprenant qu'il est essentiel de représenter l'ensemble des points de vue et que les détenteurs d'enjeux pourraient vouloir exercer un plus grand contrôle sur l'information générée, les répondants constatent cependant que l'évaluateur doit inexorablement effectuer des choix parce que l'information est trop abondante. Il est inévitable que certains groupes ou certaines personnes se sentent lésés à la fin de ce processus d'épuration.

Condition 4 : développer une argumentation cohérente

Pour la plupart des répondants (5 sur 6), la cohérence du développement constitue la pierre angulaire d'un jugement crédible parce qu'il assure sa justification, son argumentation. Dans cette perspective, l'évaluateur doit mettre en évidence le fil conducteur entre la question d'évaluation et/ou les objectifs, la méthodologie utilisée, les modalités retenues pour traiter les données, et le jugement généré. Figure 1 illustre le propos.

Cependant, les répondants notent l'absence de balise pour encadrer cette démarche, surtout en ce qui a trait au passage de l'intégration et l'harmonisation des informations à la production d'un jugement. Cette limite a nécessairement des répercussions sur la teneur de l'argumentation qui en découle. Les répondants notent que cet exercice est laissé à la discrétion des évaluateurs et que certains tirent mieux leur épingle du jeu que d'autres. Dans tous les cas, ils n'arrivent pas à expliquer pourquoi il en est ainsi.

En résumé, les répondants adoptent une position consensuelle quant à la pertinence d'un jugement crédible ainsi qu'à sa teneur. Il en est de même en ce qui concerne les conditions requises pour y parvenir.

**Figure 1**  
**Le processus argumentaire menant au jugement**



## DISCUSSION

Plusieurs considérations se dégagent des résultats obtenus.

### En ce qui concerne la méthodologie

Son importance ne ressort pas uniquement dans les témoignages. En effet, elle occupe une place importante au plan des écrits portant sur l'évaluation de programme. Si les répondants reconnaissent que la qualité de la méthodologie utilisée s'avère essentielle pour obtenir un jugement crédible, leurs propos rejoignent néanmoins ceux, plus nuancés, de Schwandt (2009) que « what constitutes credible evidence cannot be decided by method choice alone » (p. 209).

### En ce qui concerne la qualité de l'information

Il est à noter que les répondants tiennent un discours assez similaire à celui de Stake et Schwandt (2006) qui ont introduit les concepts de qualité *mesurée* (« quality as measured ») et de qualité *ressentie* (« quality as experienced »). Alors que la première est associée aux

données quantitatives et qualitatives et qu'elle permet d'assurer une légitimité scientifique à la démarche évaluative, la seconde « draws attention to both the subjective and intersubjective meanings we attach to events, personal encounters, and places » (Stake & Schwandt, 2006, p. 408). Dans la même ligne de pensée, certains répondants rapportent qu'il est aussi complexe de générer un jugement dans le cadre de l'évaluation de programme que dans les milieux médical et judiciaire. En effet, le médecin tout comme le juge ne se fie pas uniquement aux données recueillies, mais également au discours du malade, du plaignant, ou de leurs proches afin d'établir un diagnostic ou un verdict. Comme le mentionne Datta (2006), l'agencement des différents types d'information est important et il constitue un défi de taille.

#### En ce qui concerne l'implication des détenteurs d'enjeux

House (1980), Lincoln et Guba (1985), House et Howe (1999), Chen (2004), Greene (2006), ainsi que Stake et Schwandt (2006) soulignent l'importance d'une bonne communication entre les détenteurs d'enjeux et l'évaluateur. Dans le cadre de la présente étude, les répondants reconnaissent l'importance de faire participer les détenteurs d'enjeux à la démarche évaluative afin de dresser un portrait le plus fidèle à la réalité, mais ils avouent que la gestion des rapports n'est pas toujours facile. Cependant, comme dans le cas de l'argumentation, il n'existe pas, selon eux, de balise pour encadrer le dialogue.

#### En ce qui concerne le développement d'une argumentation cohérente

Comme le mentionne Smith (1995), l'évaluation de programme se compare à l'enquête criminelle qui consiste à recueillir de l'information pertinente et rigoureuse (données et évidences) et à la synthétiser pour générer un jugement qui sera fondé alors sur une argumentation solide. Les répondants verbalisent le malaise qui existe quant à l'absence de balises pour encadrer l'évaluateur dans sa démarche pour générer une argumentation solide appuyant la teneur du jugement : cela relève de l'initiative individuelle et certains évaluateurs sont meilleurs que d'autres à cet exercice. Ce constat rejoint les propos de Datta (2006) et Stern (2006) à l'effet que les praticiens ne disposent pas, à ce jour, de modalités pour synthétiser l'information recueillie afin d'élaborer une solide argumentation. Scriven avait d'ailleurs déjà noté que « the lack of any explicit justification of the synthesis procedure is the Achilles heel of the whole effort » (1994, p. 367).

Ainsi, les répondants s'entendent autant sur la teneur des conditions requises que des embûches qui se présentent. Serait-ce le recours à une connaissance plus pratique que théorique qui permet de mettre en évidence les difficultés que l'évaluateur est susceptible de rencontrer au cours du processus évaluatif. Dans tous les cas, les solutions n'apparaissent pas évidentes.

## CONCLUSION

La présente recherche visait à cerner le sens à donner au concept jugement crédible et à identifier les conditions requises pour y parvenir. Les répondants notent que, pour être crédible, un jugement devrait être solidement « argumenté, utile et explicite ». Ces critères sont explicités dans le cadre de la section sur les résultats. L'étude aura aussi permis de cerner les conditions requises pour obtenir un jugement crédible, à savoir le recours à une méthodologie rigoureuse, la production d'une information de qualité, la prise en compte de différents points de vue, et la présentation d'un développement argumentaire cohérent. Toutefois, il semble qu'en dépit de la présence de ces conditions, l'absence de balise pour encadrer la démarche évaluative, notamment en ce qui a trait au passage de l'intégration et l'harmonisation des informations à la production d'un jugement, semble être un frein significatif. Enfin, quoique nécessaire, le dialogue entre les différentes parties se transforme parfois en dialogue de sourds et l'argumentation qui soutient le jugement relève souvent plus de l'art qu'autre chose. Ainsi, en plus de dégager d'éventuelles pistes de futures recherches, la présente étude met en évidence, dans une certaine mesure, la complexité du passage entre les connaissances théoriques et pratiques.

## FORCES ET LIMITES DE L'ÉTUDE

Il est intéressant de noter qu'en générant des données, la présente étude s'appuie sur des connaissances pratiques plutôt que théoriques. De plus, la méthodologie s'appuie sur les critères établis par Patton (2002) et Stake (2010) afin d'assurer de la rigueur à l'étude. En effet, le choix des entrevues en profondeur a permis d'identifier les éléments pertinents concernant la problématique et à les documenter. De plus, la *triangulation des sources d'information* (différents points de vue et différentes organisations) ainsi que la *triangulation de l'analyse des données* (confrontation des analyses indépendantes effectuées par deux chercheurs) ont favorisé l'enrichissement du produit final.

Finalement, non seulement les entrevues ont-elles été réalisées par une personne d'expérience, qui a favorisé une conceptualisation et une théorisation émergente, mais en plus elles ont été retranscrites afin que les analyses soient effectuées à partir de verbatims. Par contre, le nombre des répondants demeure limité. De plus, on se rend compte après coup que le recours à des rapports d'évaluation comme mises en situation n'aura pas permis de mettre en évidence de façon efficace le processus d'argumentation sous-jacent au jugement. En effet, ce processus relève plus d'une démarche réflexive initiée par l'évaluateur, et ce, parfois même à son insu. Pour être capté, cela aurait nécessité qu'on interroge directement les évaluateurs, or ils ne faisaient pas partie d'aucun groupe. Par contre, ce constat a été retenu pour la prochaine phase du programme de recherche.

## RÉFÉRENCES

- Alkin, M.C., & Christie, C.A. (2004). An evaluation theory tree. Dans M.C. Alkin (Éd.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–65). Thousand Oaks, CA : Sage.
- Arens, S.A. (2006). Le jugement évaluatif et la pratique. *Mesure et Évaluation en Éducation*, 29, 45–56.
- Blais, M., & Martineau, M. (2006). L'analyse inductive générale: description d'une démarche visant à donner un sens à des données brutes. *Recherche qualitative*, 26, 1–18.
- Bossiroy, A. (2009). *Mieux comprendre la pratique évaluative : les éléments qui contribuent à la production d'un jugement fondé et argumenté*. Mémoire de maîtrise en éducation, Université du Québec à Montréal. Montréal, QC.
- Champagne, F., Contandriopoulos, A.P., Brousselle, A., Hartz, Z., & Denis, J.-L. (2009). L'évaluation dans le domaine de la santé : concepts et méthodes. Dans A. Brousselle, F. Champagne, A.P. Contandriopoulos, & Z. Harts (Éds.), *L'évaluation : concepts et méthodes* (pp. 35–56), Montréal, QC : Les Presses de l'Université de Montréal.
- Charmaz, K. (2002). Qualitative interviewing and Grounded Theory analysis. Dans J.F. Gubrium & J.A. Holstein (Éds.), *Handbook of interview research. Context & method* (pp. 675–694). Thousand Oaks, CA : Sage.
- Charmaz, K. (2005). Grounded Theory in the 21th century. Dans N.K. Denzin & Y.S. Lincoln (Éds.), *The Sage handbook of qualitative research* (3ième éd.) (pp. 507–535). Thousand Oaks, CA : Sage.

- Chen, H.-T. (2004). The roots of theory-driven evaluation. Dans M.C. Alkin (Éd.), *Evaluation roots. Tracing theorists' views and influences* (pp. 132–152). Thousand Oaks, CA : Sage.
- Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA : Sage.
- Christie, C.A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–36.
- Daigneault, P.-M. (2010). L'examen de la qualité des évaluations fédérales : une méta-évaluation réussie? *The Canadian Journal of Program Evaluation* 25(2), 191–224.
- Datta, L.-E. (2006). The practice of evaluation challenges and new directions. Dans I.F. Shaw, J.C. Greene, & M.M. Mark (Éds.), *The Sage handbook of evaluation* (pp. 419–438). Thousand Oaks, CA : Sage.
- Donaldson, S.I., & Lipsey, M.W. (2008). Roles for theory in contemporary evaluation practice: Developing practical knowledge. Dans J. Shaw, J. Greene, & M. Marks, (Éds.), *Handbook of evaluation* (pp. 56–75). Thousand Oaks, CA : Sage
- Donaldson, S.I., Christie, C.A., & Mark, M.M. (Eds.) (2009). *What counts as credible evidence in applied research and evaluation practice?* Los Angeles, CA : Sage.
- Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program evaluation. Alternative approaches and practical guidelines* (4ième éd.). Boston, MA : Pearson.
- Fish, D., & Coles, C. (1998). *Developing professional judgment in health care: Learning through the critical appreciation of practice*. Boston, MA : Butterworth-Heinemann.
- Fournier, D.M. (1995). Establishing evaluative conclusions : A distinction between general and working logic. Dans D.M. Fournier (Éd.), *New Directions for Evaluation. Reasoning in evaluation: Inferential links and leaps* (68, pp. 15–32). San Francisco, CA : Jossey-Bass.
- Fournier, D., & Smith, N.L. (1993). Clarifying the merits of argument in evaluation practice. *Evaluation and Program Planning*, 16, 315–323.

- Given, L. (2008) (Éd.). *The Sage encyclopedia of qualitative research methods*. (2, pp. 697–698). Thousand Oaks, CA : Sage.
- Greene, J.C. (2006). Evaluation, democracy, and social change. Dans J. Shaw, J. Greene, & M. Marks, (Éds.), *Handbook of evaluation* (pp. 118–160). Thousand Oaks, CA: Sage.
- Guba, E.G. (1972). The failure of educational evaluation. Dans C.H. Weiss (Éd), *Evaluating action programs: Readings in social action and education* (pp. 251–266). Boston , MA : Allyn and Bacon.
- Guba, E.G., & Lincoln, Y.S. (1981). *Effective evaluation : Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA : Jossey-Bass
- Guillemette, F. (2006). L'approche de la *Grounded Theory*; pour innover? *Recherches Qualitatives*, 26, 32–50.
- Henry, G.T. (2000). Why not use? Dans V.J. Caracelli & H. Preskill (Éds.), *The expanding scope of evaluation use. New Directions for Evaluation* (88, pp. 85–98), San Francisco, CA : Jossey-Bass.
- House, E.R. (1980). *Evaluating with validity*. Beverly Hills, CA : Sage.
- House, E.R., & Howe, K.R. (1999). *Values in evaluation and social research*. Thousand Oaks, CA : Sage.
- 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 Evaluation en Education*, 29, 27–44.
- Hurteau, M., Houle, S., & Mongiat, S. (2009). How legitimate and justified are judgements in program evaluation? *Evaluation. The International Journal of Theory, Research and Practice*, 15, 309–317.
- King, J.A. (2002). Building the evaluation capacity of a school district. Dans D.W. Compton, M. Braiserman, & S.H. Stockdill (Éds.), *The art, craft, and science of evaaution capacity building. New Directions for Evaluation* (93, pp. 63–80), San Francisco, CA : Jossey-Bass.
- Kundin, D.M. (2010). A conceptual framework for how evaluations make everyday practice decisions. *American Journal of Evaluation*, 31, 347–362.



- Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA : Sage.
- Mark, M.M., Henry, G.T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA : Jossey-Bass.
- Mathison, S.M. (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA : Sage.
- Nevo, D. (2006). Evaluation in education. Dans I.F. Shaw, J.C. Greene, & M.M. Mark (Eds.), *The Sage Handbook of Evaluation*, (pp. 441–460). London: Sage.
- Patton, M.Q. (1994). Developmental evaluation. *Evaluation Practice*, 15, 311–319.
- Patton, M.Q. (1997). *Utilization focused evaluation: The new century text*. Thousand Oaks, CA : Sage.
- Patton, M.Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA : Sage.
- Patton, M.Q. (2005). Developmental evaluation. Dans S. Mathison (Éd.), *Encyclopedia of evaluation*. Thousand Oaks, CA : Sage.
- Preskill, H., & Torres, R.T. (1998). *Evaluation inquiry for learning in organizations*. Thousand Oaks, CA : Sage.
- Preskill, H., & Torres, R.T. (2000). The learning dimension of evaluation use. Dans V.J. Caracelli & H. Preskill (Éds.), *The expanding scope of evaluation use. New directions for evaluation* (88, pp. 25–38), San Francisco, CA : Jossey-Bass.
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY : Basic Books.
- Schwandt, T.A. (2007). *Educating for intelligent belief in evaluation*. *Actes du congrès annuel de l'Association Américaine d'Évaluation*, Baltimore, MD.
- Schwandt, T.R. (2009). Toward a practical theory of evidence for evaluation. Dans S.I. Donaldson, C.A. Christie, & M.M. Mark (Éds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 197–212), Los Angeles, CA : Sage.

- Scriven, M. (1967). The methodology of evaluation. Dans R.W. Tyler, R.M. Gagné, & M. Scriven (Éds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Scriven, M. (1980). *The logic of evaluation*, Inverness, CA : Edgepress.
- Scriven, M. (1991). *Evaluation thesaurus* (4ième éd.), Newbury Park, CA : Sage.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice*, 15, 367–382.
- Scriven, M. (1995). The logic of evaluation and evaluation practice. Dans D.M. Fournier (Éd.), *New directions for evaluation. Reasoning in evaluation : Inferential links and leaps* (68, pp. 15–32), San Francisco, CA : Jossey-Bass.
- Scriven, M. (2007). Key evaluation checklist. Repéré le 29 août, 2011, de <[http://www.wmich.edu/evalctr/archive\\_checklists/kec\\_feb07.pdf](http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf)>.
- Secrétariat du Conseil du Trésor du Canada. (2004). Examen de la qualité des évaluations dans les ministères et les organismes. Repéré en 2005 de <[http://www.tbs-sct.gc.ca/eval/pubs/rev-exam\\_f.asp](http://www.tbs-sct.gc.ca/eval/pubs/rev-exam_f.asp)>.
- Smith, N.L. (1995). The influence of societal games on the methodology of evaluative inquiry. *New directions for evaluation* (68, pp. 5–14), San Francisco, CA : Jossey-Bass.
- Stake, R.E. (2004). *Standards-based & responsive evaluation*, Thousand Oaks, CA : Sage.
- Stake, R.E. (2010). *Qualitative research. Studying how things work*. New York, NY : Guilford.
- Stake, R., Migotsky, C., Davis, R., Cisneros, E.J., Depaul, G., Dunbar, C., et al. (1997). The evolving syntheses of program value. *Evaluation Practice*, 18, 89–103.
- Stake, R.E., & Schwandt, T.A. (2006). On discerning quality in evaluation. Dans I.F. Shaw, J.C. Greene, & M.M. Mark (Éds.), *The Sage handbook of evaluation* (pp. 404–418). London : Sage.
- Stern, E. (2006). Contextual challenges for evaluation practice. Dans I.F. Shaw, J.C. Greene, & M.M. Mark (Éds.), *The Sage handbook of evaluation* (pp. 292–314). Thousand Oaks, CA : Sage.

- Talmage, H. (1982). Evaluation of programs. Dans H.E. Mitzel (Éd.), *Encyclopedia of educational research* (5ième éd.), New York, NY : Free Press.
- Tourmen, C. (2009). Evaluators' decision making. The relationship between theory, practice, and experience. *American Journal of Evaluation*, 30, 7–30.
- Weiss, C.H. (1998). *Evaluation methods for studying programs and policies*. Upper Saddle River, NJ : Prentice Hall.

**Marthe Hurteau** est professeure au département d'éducation et pédagogie, Faculté des sciences de l'éducation de l'Université du Québec à Montréal (UQAM). Elle détient le titre d'évaluatrice accréditée décerné par la Société canadienne d'évaluation (SCE).

**Pierre Valois** est professeur en sciences éducatives, Faculté des lettres, des sciences humaines, des arts et des sciences de l'éducation de l'Université du Luxembourg

**Amélie Bossiroy** est diplômée du programme de maîtrise en éducation de l'Université du Québec à Montréal. Au cours de ses études elle s'est spécialisée en mesure et évaluation, et plus spécifiquement, en évaluation de programme.